SAIC -88/1898

LICENSING SUPPORT SYSTEM CONCEPT FEASIBILITY ANALYSIS

SCIENCE APPLICATIONS INTERNATIONAL CORPORATION

July 10, 1988

LSS SYSTEM CONCEPT FEASIBILITY ANALYSIS

TABLE OF CONTENTS

Preface

This report fulfills the contract milestone for a Systems Concept Feasibility Analysis as required by the contract between the Department of Energy Office of Civilian Radioactive Waste Management (OCRWM) and Science Applications International Corporation (SAIC) for the Licensing Support System (LSS). The analysis effort has been an ongoing process as a part of the preparation of the series of four reports which have been prepared for OCRWM submittal to the Office of Management and Budget over the past eight months. Therefore these reports, which constitute the body of work performed on the Systems Concept Feasibility Analysis, are incorporated as attachments to this report.

Errata

The following errata are provided for the Attachments:

<u>Attachment B. Preliminary Data Scope Analysis</u>

Page 45, Item (2), Last line, Reference should read DOE, 1988d

Page 46, Table 8, Replace with Table in Appendix B of the Conceptual Design Analysis

<u>Attachment C. Conceptual Design Analysis</u>

Page ii, Item (5), Fifth word should be "headers"

Page 15, Sect. 2.2.2, Lines 4,5, Delete "non-textual"

Page 40, Sect. 3.2.3.1, Line 2, First word should be "field"

<u>Attachment D. Benefit-Cost Analysis</u>

Page iii, Third line from bottom should read "Telecommunications ... 2%"

Page 27, Item (3), Sixth line should read "Telecommunications ... 2%"

1.0 INTRODUCTION

As a part of the initial Licensing Support System (LSS) development effort, Science Applications International Corporation (SAIC) has prepared a series of four reports for the DOE Office of Civilian Radioactive Waste Management (OCRWM). These reports have in turn been submitted to the Office of Management and Budget by OCRWM to document the basis for the LSS design. The reports have covered the entire conceptual design process, from identification of needs and requirements to analysis of benefits and cost. The reports constitute the basis of the System Concept Feasibility Analysis and have been assembled here.

1.1 Purpose and Scope

The LSS System Concept Feasibility Analysis was originally intended to provide DOE with the assurances from the LSS contractor that the conceptual design as contemplated in the Functional Requirements Document is both technically achievable and can be attained within reasonable costs. In light of the fact that the High Level Waste LSS Advisory Committee was determining LSS requirements as part of the negotiated rulemaking process, it became necessary to revise the process somewhat. SAIC therefore performed a complete conceptual design: integrating the requirements from the negotiated rulemaking process with needs identified from user interviews, determining the scope of the information to be stored, developing a conceptual design with reasonable variations, and analyzing the costs and benefits of the designs. During the course of developing the conceptual design and the benefit-cost analysis, the feasibility of the design was constantly being evaluated. This report will focus on this feasibility aspect of the design process.

This report has been organized into three major sections. Section 1 provides an introduction to the report. Section 2 is a summary of the information contained in the four previous reports with specific information on the aspects of concept feasibility that are explicitly or implicitly contained in those reports. Section 3 provides the conclusions, and the four reports comprising the basis of the analysis are included as attachments.

1.2 Background

The development phase of a system integration contract, either explicitly or implicitly, includes an analysis of the feasibility of the total system. Feasibility in this context generally means a reasonable assurance that the system conceptual design is capable of being produced without unreasonable technical risk or cost, and that the system will perform according to the requirements and specifications.

Feasibility can be analyzed and demonstrated in a number of ways. The most common way, and the way that has been chosen here, is to compare the requirements and functionality of the system (in this case the LSS) with similar systems which have been produced, demonstrated, or are being marketed by vendors. In this manner it can be shown that (at least on a function by function basis) the LSS Conceptual Design is based on proven components and is not dependent upon research or unproven technology developments. Two further questions must be considered, however, to complete the analysis. First, does the LSS concept require that any component, whether it be hardware or software, operate near or beyond their limits of performance that have thus far been successfully demonstrated? Second, can the feasibility of all components and functions, working together, be determined?

The feasibility analysis for the LSS is based on a division of the total system and functionality into subsystems which have been separately analyzed. Then by minimizing the interdependence of the subsystems on each other, the question of total system performance can be addressed in the conceptual design.

2.0 CONCEPT FEASIBILITY

The conceptual design analysis which has been performed for the LSS consisted of four steps, each of which culminated in a report documenting the results. In this section the reports will be summarized, paying particular attention to the feasibility aspects which are inherent in the process. Additional detail in any of the four steps can be found in the reports themselves which are included as attachments.

2.1 Preliminary Needs Analysis

Several background studies were performed prior to this contract which indicated the need for a sophisticated computer system to support the High Level Waste repository licensing process. General estimates of the size and variety of the document data base and the number and geographic distribution of users suggested an advanced computer system that will make use of state-of-the-art search techniques. However these studies lacked the necessary detail to develop reliable system specifications. Thus a focused study of user needs was determined necessary to ascertain specific requirements.

The process of determining the user needs consisted of identifying user categories and interviewing representative members of each. This was particularly challenging since many of the job functions which will use the LSS have not yet been defined. In many cases persons able to anticipate the representative LSS usage were relied upon. The results of the interview process, combined with the system requirements defined in the draft 10CFR Part 2 Subpart J, form the basis for the LSS system user requirements.

The needs analysis focused on a determination of system performance requirements and information access features which the users determined would be necessary to aid them in performing their assignments. Specific aspects included user session characteristics, geographic distribution and number of users, required response time, output requirements, access techniques, features of full-text and structured index searching, and system-user interface.

The conclusions from this analysis are detailed in the Preliminary Needs Analysis (Attachment A), and summarized below:

1) Some information should be stored in the form of headers for every document and full-text for many, if not all documents.

2) Capability should be provided for efficient and accurate data retrieval using a variety of methods including structured index searching on detailed headers and full text search on both headers and document text.

3) Search aids including a thesaurus, Boolian logic, and proximity searches should be available.

4) Output to the user should include hard copy of documents and reports as well as screen output of ASCII text and search results.

5) The LSS should be designed for ease of use with minimum training, contain built-in help functions, and should provide assistance as needed, either through an expert system or on-call assistance.

6) Electronic mail is required for the communication of hearing-related documents and notifications.

7) The data base should be secure and contain minimum duplication.

8) The number of users will probably exceed 350 at peak demand and will primarily be located in Washington D.C., Las Vegas, Nevada, and San Antonio, Texas

9) Electronic (bit-mapped) images of documents may be required for some records, particularly graphic oriented material.

Since the publication of this report, the draft version of 10 CFR Part 2, Subpart J has been negotiated. Assuming that this rule survives the comment process in its present form (as is likely), the LSS will be required to store all textual material in a form in which documents can be full-text searched.

The question of feasibility did not enter into the needs determination directly. However, in the interview process, most users had had previous experience with computer-based information systems. Thus most of their needs were expressed in terms of capabilities which are known to exist in current systems. Nevertheless, that is not to say that all of the expressed needs are technically or economically feasible in the specific context of the LSS.

2.2 Preliminary Data Scope Analysis

This analysis was performed specifically to estimate the size of the data base required to be stored and searched in the LSS. Combined with the needs analysis identified above, the full system requirements can be established. The analysis consisted of a review of current document files which are considered to be applicable to the LSS scope of purview, such as the DOE headquarters OCRWM records management data base, the NNWSI (Las Vegas) records management data base, and the NRC data base on nuclear waste management. These were reviewed for applicability to the LSS, duplication of documents, and rate of growth in order to estimate the current and projected size of the data base in 1990 (the expected date to begin data capture). Finally the growth of the data base from 1990 to 2009 was projected using a profile of programmatic activity developed from major repository program milestones. An important conclusion of this analysis was that there is (and will continue to be) appreciable unavoidable uncertainty in making estimates of the LSS data base size under any given set of conditions, overlain by additional uncertainty about the stability of these conditions. Since this process involves a significant level of uncertainty, a range of values were calculated providing both a low and a high estimate of the number of pages which would be candidates for inclusion in the LSS. The results of this analysis projected the candidate data base size to be:

	Cumulative Pages at Year End	
Date	Low Estimate	High Estimate
1990	9,304,000	11,885,000
1998	21,404,000	27,921,000
2009	32,191,000	42,216,000

While the negotiated rulemaking process resulted in a topical definition of documents to be included in the LSS after this estimate was complete, an analysis of the topical definitions did not indicate any significant difference from the methodology used to create this estimate. Thus this remains to date the best projection of the data scope for the LSS.

The concept of feasibility did not enter into this phase of the work in order that the estimates not be biased with concerns over size of the system.

2.3 Conceptual Design Analysis

The Preliminary Needs Analysis, Preliminary Data Scope Analysis, and the results of the Negotiated Rulemaking process defined the requirements of an automated information storage and retrieval system. A conceptual design was formulated which meets the requirements and contains the following primary features:

1) Headers and searchable full text of all documents,
2) Bit-mapped images of all documents for hardcopy reproduction and distribution as well as on-line display and local printing,
3) Centralized text and image storage,
4) Multiple geographically distributed capture systems,
5) Workstations capable of displaying text and images, and support for user-owned workstations for text display,
6) Retrieval through structured index searching of cataloged information and software full-text searching of document text, and
7) Electronic mail.

The primary application of the concept of feasibility entered into this phase of the design. As the design concept evolved, it was reviewed from the standpoint of the ability of current or near-term technology to accomplish the required need, and the overall system capability to perform the total mission. This concern resulted in a division of the total LSS functionality into systems which can, for the most part, be evaluated individually. The major systems of the LSS conceptual design are:

Capture:	Receipt and preparation of the data for storage
Search:	Storage and retrieval of text and structured data
Image:	Storage and retrieval of page images
Communications:	Transmission of data with the user and among systems
Workstations:	User interface to the LSS

By subdividing the required LSS functionality and assigning functions to the various systems, the feasibility of each system could better be determined.

During operation the LSS data base will continuously increase in size as new documentation is prepared and distributed by all of the concerned parties to the licensing process. Nevertheless it is not practical to consider sizing the system for a data base which is larger than can be anticipated in more than ten years of operation, because computer-based information systems typically are replaced due to obsolescence in five to seven years. The LSS Conceptual Design was sized to contain the data base which is anticipated in 1998 (based on the "high estimate" of pages). This date was chosen as the maximum in which a computer system can be anticipated to be supported (8-9 years from purchase), and also because it encompasses the completion of a major milestone in the repository licensing process, i.e. issuance of the construction authorization. Thus the LSS design is not dependent on future technological developments in order to fulfill its primary mission to support the first licensing step.

2.3.1 Capture System

The function of the capture system is to receive documents from the submitter and prepare them for loading onto the LSS. This process involves scanning the document to obtain the electronic image, conversion of the image to ASCII text (if the text is not already available from word processing), entering of the bibliographic and subject data into a structured data base, microfilming for archive records, and preparing the resulting data for loading into the LSS. Each of these functions is currently being performed in production environments such that individually the feasibility has been proven. An analysis of the size of the backlog of candidate documents and the time allowed to load the backlog (specified in the draft 10 CFR Part 2, Subpart J) resulted in a required sustained processing load of 18,000 pages/day. Since current proven data capture capability is approximately 3,000 pages/day, it was decided to further subdivide the Capture Subsystem into "Capture Stations", each of which can process approximately 3,000 pages/day. In this manner the feasibility of the capture process can be assured. Improvements in component capabilities, particularly in the speed and accuracy of the optical character recognition (OCR) process, will further reduce the labor requirements and improve the capabilities of the Capture System.

2.3.2 Search System

The Search System represents the most difficult functional requirements to evaluate. This system must index and store all of the text as well as the structured data base containing the bibliographic and subject data (headers). The concept itself is not difficult to accomplish as commercially available software is available to perform this function. The problem lies in the size of the data base to be captured, the number of simultaneous users to be serviced, and achieving acceptable response times. In order to determine if the requirements on the computer system were within the realm of possibility, and in order to size the computer for cost estimates, a detailed computer model was developed. Inputs to the model included the expected transactions to be performed by the users in terms of text searches, structured data base searches, requests for reports, etc.; results included the required processing power to service the users and the expected response time. Comparing the results to currently available computer offerings, both in terms of mainframes and clustered mini-computers, revealed that the Search System would be within the capabilities of current computer systems. Additionally, the analysis revealed that a "single" data base of the size expected for LSS was not feasible, and that the data base must be partitioned. All of the commercial and government full text data bases of a size comparable to LSS utilize this concept. Feasibility of the Search System can therefore be determined provided the assumptions used in the analysis, particularly user query profiles, are correct. One of the major aspects of the LSS prototype testing phase will be to more closely define user query profiles and complexity.

2.3.3 Image System

By separating the image retrieval function from the Search System, the primary function of the Image System becomes one of a "file server" to retrieve and transmit images (via the Communications System) to the user workstation or a high-speed printer for hard copy production. As such, the functionality of this system is within the capabilities of commercially available hardware and software for image file control in terms of the expected usage of images in the LSS. That is to say, the Image System is designed to support limited on-line access to images by some users and access for background and overnight printing. Continuous "browsing" of images by most users would result in long image access times. Again the question of data base size is a problem area, for while vendors advertise the capability to tie multiple optical disk jukeboxes to a file server, the size of the anticipated LSS image data base (1,400 gigabytes by 1998) is larger than any system currently on-line.

2.3.4 Communications System

The conceptual design envisions that the Capture System will provide data to the Search and Image Systems by manually transmitting magnetic media and/or optical disks via mail or delivery service. Therefore the primary function of the Communication System is to link the user workstations to the

LSS data bases and electronic mail capability. The communications design is primarily a compatibility, sizing and optimization effort to ensure adequate system response times to the user. The types of communications devices employed are dependent upon the geographic location, number of users at that location, required system functionality, and the characteristics of the particular host computer and workstation. The devices will be combinations of local area networks (LAN), LAN bridges, multiplexors, modems, communication processors, high-speed leased lines, and standard telephone switched networks. All of these devices are commonly used data communications equipment in computer networks and therefore do not present a question of feasibility. In order to evaluate the overall system communications requirements and performance, a traffic analysis model was constructed and analyzed for typical user scenarios. In this manner the feasibility of the communications system as a whole was evaluated.

On a local functional level, the only question of feasibility pertains to the display of images to terminals which are connected to the LSS via standard phone lines and modems (remote locations with few users). The relative slower speeds of the devices mean that the display of images at these locations is not feasible (i.e. image display time is too slow). If images must indeed be displayed at these locations, higher speed communications devices must be employed. It should be noted that the requirements analysis did not indicate that images need to be displayed to all users.

2.3.5 Workstations

The basic functions of the user workstation include connectivity to the Search System, file transfer, and local printing. All of these functions are well within the capabilities of micro- ("personal") computer devices; therefore the basic or Level 1 workstation will be based on this concept. For those users which will require the display of electronic images at the workstation, the Level 2 workstation will be a micro-computer with the addition of a high-resolution graphics monitor and an image decompression interface board. This concept is already in use in many applications, thus the feasibility is demonstrated.

2.4 Benefit-Cost Analysis

The conceptual design analysis included the identification of several alternative approaches to providing the required functionality. The primary purpose of the analysis of benefits and costs was to investigate these alternatives from the standpoint of relative costs and relative benefit to the user. At the same time, the absolute cost of the various alternatives were calculated for a 10 year life-cycle.

The functional capabilities of the Base Conceptual Design were covered in Section 2.3. The variants to the base differed in the following ways:

I. Two search and image systems replicating the data base, one in

Washington D.C. and one in Las Vegas

II. Hardware full text search, rather than software

III. No workstations capable of displaying images

IV. Microform digitization rather than optical disk storage of images

V. Off-line microform printing rather than on-line bit-mapped image system

VI. Re-keying text rather than text conversion from scanned bit-map

VII. Combination of Variants III, V, and VI above.

The major results of the study were that the overall 10 year life cycle costs of the various alternatives were between $192 million and $236 million in 1988 dollars. The predominant factor in the cost is associated with the data capture, accounting for approximately 62 percent of the total. In other terms, labor was by far the largest cost category, accounting for approximately two-thirds of the cost, with hardware purchase and maintenance being second at nearly 20 percent.

From the standpoint of feasibility, the Benefit-Cost analysis provided the information from which one can judge the relative merits of the alternatives. The conclusion reached in the analysis was that the Base Conceptual Design provided the best combination of cost and benefit and is therefore the preferred design approach. Since the Base Conceptual Design was also chosen as the alternative which presented the least risk in development, this conclusion did not affect the feasibility of the overall project.

In addition to the relative merits and cost of the alternatives this analysis also provided the results of a 10 year life-cycle costing of the design and operation of the LSS from which one could judge the overall financial feasibility of the project.

3.0 CONCLUSIONS

The concept of feasibility has been incorporated into the conceptual design process of the LSS as documented in the attached series of reports. The overall functional requirements of the LSS are complicated and varied. By subdividing the LSS into systems, assigning functions to those systems, and then evaluating the feasibility of those systems operating individually and in concert, the feasibility of the total LSS can be demonstrated.

In each case, feasibility is demonstrated by comparison of the LSS with similar systems which have been produced, demonstrated, or are being marketed by vendors, thus demonstrating that the LSS Conceptual Design is not dependent upon research or unproven technology. Further, the LSS is not contemplated to utilize hardware or software components beyond their limits of design capability. By providing maximum independence of subsystems, there is reasonable assurance that the subsystems as a whole will function as a complete LSS.

While the LSS is not dependent upon technological advances which are expected in this field, those advances can only provide further conservatism in the feasibility of operation.

The largest unknowns in the assessment of feasibility relate not to the technology, but rather to the assumptions in user performance. For example, relatively small changes in query complexity can result in large changes in computational requirements. Thus while feasibility can be determined for the assumed user profiles used in this analysis, if the prototype results differ significantly from these assumptions, the question of feasibility needs to be re-examined.

LICENSING SUPPORT SYSTEM
PRELIMINARY NEEDS ANALYSIS

SCIENCE APPLICATIONS INTERNATIONAL CORPORATION

SAIC™

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Preface

This is the first in a series of four reports on the Licensing Support System (LSS) prepared by the DOE Office of Civilian Radioactive Waste Management (OCRWM) for the Office of Management and Budget (OMB). The LSS is an information management system intended to support the needs of all the parties involved in repository licensing, including the Department of Energy (DOE) and the Nuclear Regulatory Commission (NRC). The reports in this series are:

Preliminary Needs Analysis

Preliminary Data Scope Analysis

Conceptual Design Analysis

Benefit-Cost Analysis

The Preliminary Needs Analysis, presented in this report, and the Preliminary Data Scope constitute the system requirements basis for developing a conceptual LSS design, which will be presented in the third report. The Benefit-Cost Analysis evaluates alternatives within this conceptual design. These four reports, and subsequent refinements, are intended to provide the basis for determining the LSS design specifications.

1.0 INTRODUCTION

1.1 Purpose and Scope

The LSS computer system design being developed by OCRWM to support the requirements of all parties in the repository licensing process (including NRC and DOE) will be based on a detailed set of system specifications. These specifications will be derived from statutory, programmatic and user requirements. This Preliminary Needs Analysis, together with the Preliminary Data Scope Analysis (next in this series of reports), is a first effort under the LSS Design and Implementation Contract toward developing a sound requirements foundation for subsequent design work. These reports are preliminary. Further refinements must be made before requirements can be specified in sufficient detail to provide a basis for suitably specific system specifications.

This preliminary analysis of the LSS requirements has been divided into a "needs" and a "data scope" portion only for project management and scheduling reasons. The Preliminary Data Scope Analysis will address all issues concerning the content and size of the LSS data base; providing the requirements basis for data acquisition, cataloging and storage sizing specifications. This report addresses all other requirements for the LSS.

According to the definition in the LSS System Development RFP (DOE, 1987), the LSS consists of both computer subsystems and non-computer archives. This study addresses only the computer subsystems, focusing on the Access Subsystems.

After providing background on previous LSS-related work, this report summarizes the findings from previous examinations of needs and describes a number of other requirements that have an impact on the LSS. The results of interviews conducted for this report are then described and analyzed. The final section of the report brings all of the key findings together and describes how these needs analyses will continue to be refined and utilized in on-going design activities.

1.2 Background

The Nuclear Waste Policy Act (NWPA) of 1982 authorizes the siting and construction of the nation's first permanent repository for spent fuel rods and high-level nuclear waste. Under the law, the US Department of Energy (DOE) is responsible for siting and construction, the Nuclear Regulatory Commission (NRC) for licensing, and the nuclear utilities for providing the funding. The NRC is allowed three years in which to consider the application for a license to authorize repository construction.

Shortly after the passage of the NWPA, the NRC became convinced that it would be impossible, under existing procedures for the conduct of high-level nuclear waste repository licensing proceedings, to complete the proceedings within the three years allowed by the Act. The principal reason for this conclusion lay in two aspects of document discovery, the process by which parties and intervenors identify and exchange documents relevant to the issues pertaining to repository licensing. The first aspect was document volume: a typical large case before an Atomic Safety and Licensing Board (ASLB) of the NRC generates a record of 10,000 documents. The number of documents generated in connection with the repository design was expected to be 300 to 1000 times greater, depending upon whether discovery was directed to three sites or one. The second aspect was delay inherent in the discovery process itself. (Jordan, 1986) Under existing discovery rules, a document production request in a large case can require 12 to 18 months of manual effort. Large file rooms have to be established by each party and time has to be provided to manually sort and select records at the site of production. With multiple well-funded parties this means extensive travel, scheduling, review, and motion practice. (Olmstead, 1986) To meet a three-year licensing timeframe, this time-consuming activity must be substantially reduced.

Based on the above concerns, in 1984 the NRC's Division of Waste Management initiated efforts to further scope the problem. The regulatory staff met with DOE (including the three field offices) and the affected States and Indian Tribes. It was found that there was general agreement on the need for a common computerized system to handle the massive volume of documents and accommodate a three-year licensing review. The staff also visited agencies that currently use state-of-the-art computerized storage and retrieval systems, such as the Library of Congress and the Patent Office. However, it was found that none of the existing state-of-the-art systems would fully meet NRC's needs.

By 1985, the NRC became persuaded of a two-part solution to this apparent impasse:

- o A computerized information storage and retrieval (IS&R) system that would serve as the sole basis of discovery, processing and making available all relevant documents in advance of the licensing hearings

- o A rule, adopted by all parties, that would with a few exceptions limit discovery to what was contained in this IS&R system

In addition, removing as much document discovery as possible from the proceedings, through modifications to the NRC Rules of Practice was proposed (10 CFR 2).

This last proposal reflected concern about the second aspect of document discovery mentioned above: it did little good to create a means for rapid retrieval of information and documents in discovery if the means consumed a great deal of time during the license review period. Of the sources of delay, perhaps the most detrimental to the three-year licensing time frame is the motions practice, the filing of and response to legal motions. NRC estimated that 40 percent of a typical hearing before an ASLB was consumed in motions practice. The NRC therefore proposed that motions and responses should be filed to an ASLB computer by phone lines; in other words, by electronic mail.

With the electronic mail concept, the NRC made two additional proposals. First, it would save a great deal of time if all parties would agree to supply their documents electronically; more specifically, in a format that could be used by the IS&R system software. Documents created before a certain date could be processed into an interim IS&R data base by traditional means (i.e., by manual cataloging and abstracting) and after that date the documents could be provided in electronic format or be made machine-readable by rekeying. Under this proposal, reading the full text of documents into the data base by optical character recognition (OCR) methods could ultimately eliminate the time-consuming traditional process of document acquisition, backlogging, coding, and keying.

The second additional proposal was to store electronic images of the documents on an optical disk storage (ODS) system. When a user of the full-text IS&R system located documents germane to a query, he could transmit the documents from optical disk over telecommunications lines to his remote location. This would minimize the time spent in locating, copying and mailing the physical documents.

These additional proposals were adopted within NRC (Browning, 1985) and became the basis of two extremely relevant documents. The first of these was an agreement between NRC and DOE that the IS&R system, called the Licensing Support System (LSS), would be designed around full-text and image storage. (NRC/DOE, 1986). The second proposal appeared as notices in the Federal Register (Federal Register, 1986, 1987). These stated NRC's intent to convene a panel of interested parties to negotiate a rule on discovery for the repository licensing proceeding. In addition, these proposals became the basis of an internal NRC pilot project to test the feasibility of such concepts.

Even before the NWPA was passed, the Project Management Division of Battelle Memorial Institutes had begun generating a data base from documents created at the OCRWM Salt Repository Program Office. This data base, called the Automated Records System (ARS), was designed to meet the bibliographic retrieval needs of OCRWM scientists and engineers. It took the form of traditional surrogate records created through cataloging and abstracting. Similar data bases were later established by the Basalt Waste Investigation Program and OCRWM Headquarters and were standardized to the same software and the same data base structure.

Pursuant to the Federal Register notice mentioned above, the first meeting of the Negotiated Rulemaking Advisory Committee (NRAC), which mainly comprises attorneys representing the licensing parties and intervenors, was held in September 1987. The meeting was convened by NRC under the facilitation of the Conservation Foundation. To date, the substantive agenda for the Committee has been 20 articles in a Position Paper which has been adopted by all interested elements of NRC. While many of the articles relate to legal questions of discovery, several have served as the basis for Committee consensus on design of the LSS.

The NRAC has not yet addressed several questions on its agenda that are of vital importance to use of the LSS for discovery.

1) Should discovery, and hence the LSS, embrace documents from

sources other than the parties (i.e., other than DOE and NRC)?

2) How will it be possible to ensure that all parties' relevant documents are submitted to the LSS?

3) What are the standards of performance with which the LSS must comply?

When it issued a Request For Proposal (RFP) for the LSS in March 1987 (DOE, 1987), OCRWM was unable to specify in detail the characteristics of the system that would be required for discovery. The Committee had not yet been convened. However, it became clear that the repository licensing proceedings will be in the nature of litigation (Jordan, 1986); therefore, the LSS must possess certain characteristics not ordinarily required of an IS&R system. Administrative staff indicate that they are more interested in tracking compliance with regulations than in retrieving individual documents on specific topics.

From all this it is clear that the LSS must:

1) Serve as the sole basis for expedited document discovery,

2) Provide access to licensing information so that all parties' legal counsel and their experts can address the grounds for repository licensing decisions and determine the soundness of technical work,

3) Provide an automated library of reports and other bibliographic materials of use to OCRWM and NRC technical staff in conducting their work on licensing document development and review,

4) Serve as a mechanism for tracking OCRWM compliance with repository licensing regulations.

2.0 SOURCES OF INFORMATION ON LSS NEEDS

The needs analysis presented in this report has considered four principal sources of information:

- o The results of early needs assessments for the LSS and for similar systems
- o The current status of the negotiated rulemaking process
- o The requirements applicable to LSS that result from other than user needs (such as institutional or legal requirements)
- o The results of a survey of potential LSS users

This section presents the requirements on LSS identified in each of these four sources.

2.1 Previous Examinations of Needs

2.1.1 Requirements Definition for ARS

The Records and Information System (RIS) is a forerunner of the Automated Records System (ARS) at the OCRWM Salt Repository Project. The requirements definition for RIS began in 1979. At that time, three steps were taken to determine the needs the system was to serve:

1) Survey of Project staff

2) Survey of nuclear utilities, to profit from their lessons learned in similar projects

3) Interview visit to NRC to learn about their system's strengths and weaknesses

Survey of Project Staff

A survey of the staff of the OCRWM Salt Repository Project was conducted to determine what they thought they needed from a records management system. In summary, they asked for the following capabilities:

- o A way to account for the documents, so that none were lost. This was the first requirement for a future system to support licensing.
- o A way to locate documents and retrieve them quickly.
- o A system to replace paper files (e.g., by microfilm).

- A system that would permit rapid retrieval by subject, originator, recipient, date created and other related parameters.

- A method to track commitments to DOE.

Survey of Nuclear Utilities

In 1979, visits were made to ten nuclear utilities to learn about their records management needs and about their existing systems. The specific questions asked were:

- Was a manual or a computerized record management system in use?

- If computerized, what hardware and software were in use?

- What problems were encountered and how did they solve them?

- What did they like and not like about their systems?

- Where was records management responsibility originally located in the organization and how was the records management function organized?

- Who were the users of their system?

The utility systems were designed as inventories of the records (i.e., they assigned a number to each document and then listed them in chronological order). Only one of the ten sites used a manual system. It was extremely labor intensive and had a large backlog of documents to process.

There were many variations in the hardware and software. In general, the records management function used the equipment available in the company, which was also used by other groups. One of the major complaints across the survey was the lack of a dedicated computer, which resulted in long delays.

Most utilities reported that records management was not a high priority with their companies. There were instances in which lack of attention resulted in costly retrofits or reprocessing of records. The lack of subject search capability was also a frequent complaint.

Records management was usually located in the administrative function of the organization and was sometimes linked with the library. The mail room and control of incoming and outgoing correspondence was almost always part of records management. Rarely did the records manager report to a vice-president or higher official. Drawing control was usually a function of records management and was very active for those plants with ongoing construction projects. Users were most often the records management staff searching for known items.

The most frequently offered advice to the survey team was to get a dedicated computer and to build subject retrieval capability into the system.

Visit to NRC

The NRC correspondence and document processing functions operational in 1979 were microfiche-based and did not have subject search capability. The microfiche was retrieved by a cumbersome mechanical process. The contractor who operated the system recommended that Battelle develop a BASIS application for the operation, so that they could add a structured index retrieval capability.

Requirements for RIS

Based on information collected in the two surveys, it was decided that the RIS would be the basic records management system for the Salt Repository Project, but would also include technical literature on high-level nuclear waste disposal. It would be designed to permit searching capability of both project records and this technical literature.

The Battelle Computer Center designed the software application and CDC computers were used. The system design team worked with the records management staff so that the requirements would be understood and supported in the design. The manager of the design team is still responsible for system upgrades and modifications today. This continuity has been a valuable resource, and a critical one facilitating the ability to migrate the system from BASIS to DM. This system maintains a continuing focus on the future because of the very long-term nature of the nuclear waste program and the rapid advances in computer technology. It was understood from the beginning that a system could not be designed for a one-time use and then discarded when the software became obsolete or was superseded.

In 1986 it was determined that an integrated relational data base environment would better accommodate the needs of the RIS users. The data bases were moved from the BASIS applications to DM, which provided the relational capability. Following this transformation, the system was renamed the Automated Records System (ARS).

The documents in the RIS/ARS are cataloged for and are retrievable via structured indices by:

Date of creation	Title
Accession number	Abstract
Microfilm number	Descriptors (Keywords)
Originator/Author	Attachments
Receiver	Project ID

2.1.2 LIS Requirements Study

The Licensing Information System (LIS) was an early name for the concept that is essentially the same as the LSS. In 1985 Roy F. Weston Inc. performed a study for OCRWM (Weston, 1985) of the LIS requirements. This report provides a rich source of information on expectations for the LSS as it is being defined in 1988.

LIS Objectives

The objectives of the LIS were that a) it must serve OCRWM's current needs, b) be configured for growth and flexibility, and c) use existing information management systems in place at OCRWM Headquarters and the Project offices to the greatest extent practical.

Requirements Survey

Weston interviewed various program participants and information system specialists including OCRWM Headquarters staff, particularly those within the Office of Geologic Repositories (OGR). Licensing engineers at OCRWM and Weston were interviewed, and briefings and workshops were held with the Project offices. Meetings also were held with two nuclear facilities.

Several commercial data systems were examined in addition to the existing systems at the Project offices, including:

- Corps of Engineers Environmental Legislative Data System
- LEXIS
- JURIS.

Requirements for LIS

The requirements of the repository licensing process on the LIS include:

- The system must be interactive and comprehensive with regard to licensing information (i.e., it must be a useful tool for all participants in the licensing process).

- The system must provide rapid access to the information regardless of the user's geographic location and the geographic location of the computer system containing the information.

- The LIS record must be durable and extend for 60 to 90 years, until repository closure.

The Weston report identified nine specific requirements LIS should support:

1) Provide a comprehensive reference source of all regulations and other regulatory guidance documents applicable to the repository licensing process.

2) Establish and maintain a living licensing schedule network.

3) Identify and track all issues related to regulatory compliance, the work plans and actions directed at their resolution, and their outcome.

4) Record and track all commitments and resulting actions.

5) Document the preparation and modification of key documents needed for regulatory compliance.

6) Provide a comprehensive reference source (archive) of all information produced or captured by OCRWM which may have a bearing on regulatory compliance.

7) Maintain the confidentiality of any information which must have such protection.

8) Provide long-term storage and access to all program information which may have a bearing on regulatory compliance.

9) Provide rapid search and (if possible and practicable) full-text storage, searching and retrieval for regulatory compliance information.

System Scope

The conceptualization of this first large-scale, program-wide information system involved both a strategy toward meeting the extraordinary information needs and records management requirements of the program and a tool to be used by OCRWM during the licensing process.

As a strategy, the LIS requirements represent acknowledgment of the need for:

- A comprehensive structure to licensing data capture and management,
- An aggressive effort to identify, document and integrate the activities of OCRWM regulatory, siting, and design programs,
- An integration of the repository program's information policy.

During the interviews, no clear definition of licensing data versus program data emerged. DOE Headquarters Office of Geologic Repositories staff interviewed often suggested that any program information could be called into question under certain scenarios of the licensing process. DOE Headquarters Office of General Council staff held the view that to ignore a given segment of information as non-licensing would be counterproductive and that all information on the program will be subject to discovery during the ASLB hearings during licensing.

Full-Text Storage and Searching

The Weston report addressed full-text capture, storage and searching requirements for the LIS and cautioned that, although the technology is available, it is new and expensive. They state that "extreme care and thought needs to be exercised in deciding which regulatory compliance information should be maintained on-line in full text and in loading these documents into the system." (Weston, 1985)

Constraints

In this study the following constraints were determined to apply to the LIS:

- Existing systems and equipment should be built upon and utilized to the extent practicable;
- Users and project personnel are geographically scattered, making telecommunications networking difficult;
- The system must serve the program needs for 60 to 90 years;
- A lack of common hardware and software exists across the Project offices and DOE Headquarters; and
- There is no definition of licensing information to be included in the system (i.e., is all program information subject to discovery?).

2.1.3 Discovery Requirements Study

In 1986, John S. Jordan & Associates addressed the problem of a three year license application in a report (Jordan, 1986) to the NRC. Specifically, the report considered whether a licensing information management system might be a means of expediting discovery and providing full access to the documentation well in advance of the licensing proceeding. An examination of the legal requirements, and the problems posed by the requirements, was provided along with proposed solutions. The report described the components, characteristics, and configurations of an IS&R system that will satisfy the requirements of and implement the solutions to the legal problems.

The report primarily addresses issues related to the material to be included in an IS&R system for repository licensing support, which will be extensively discussed in the forthcoming Preliminary Data Scope Analysis report. However it also addresses five requirements issues within the scope of the present needs analysis.

1) Security. If documents that are proprietary, classified, subject to attorney-client privilege, or otherwise exempt from discovery are entered into the IS&R system, the system must have the ability to assure restricted access to specifically identified information.

2) Retrievability of information in the system must be extremely high, at or near the state of the art, for it to be acceptable for legal discovery.

3) Recall is the proportion of documents found in a search, relative to the total number of documents in the data base which meet a search criterion. This is a measure of the amount of relevant material missed in a search. The recall of the system must be as high as possible and the cost of maintaining a high degree of recall must be independent of data base size for the system to practically meet the needs of discovery.

4) Precision is the proportion of truly relevant documents found in a search, relative to the total number of documents found. This is a measure of the amount of inappropriate material returned during a search. To be useful for discovery, the precision of the system must be extremely high.

5) Speed. The size of the data base dictates that for the system to be of practical use during discovery, the system must make use of state-of-the-art technology to maximize access and minimize delay time in identifying material sought.

2.1.4 NNWSI Bridge Program

The OCRWM Nevada Nuclear Waste Storage Investigation (NNWSI) Project has developed an information management system (IMS) to support the project in areas of project management, licensing, long-term record storage, and dissemination of public information. It is envisioned that the IMS will either supply information to the LSS or more directly be the repository of project-specific information for the LSS. The general functional requirements of the IMS were stated in the Systems Concepts Evaluation Report (SAIC, 1986) and are summarized as follows:

1) Assist DOE in management of the project by providing a uniform set of project information; establishing correlation among issues, comments, regulations, work activities, commitments, documents and other key parameters; alerting DOE to unresolved compliance issues; allowing expeditious response to requests; and providing a mechanism to identify potentially contentious issues.

2) Help expedite NRC review of licensing submittals by reducing reliance on discovery through early public access to documentation; limiting the issues in contention by tracking and documenting issues; providing complete defensible submittals; and supporting timely preparation of hearing material.

3) Provide secure storage of records for extended periods and ability to retrieve documents by organizing and maintaining for retrieval all data for license amendments; providing method for controlling, storing, and retrieving records in accordance with 10 CFR 60 and 10 CFR 50 Appendix B.

4) Meet NWPA requirements for information, consultation, and cooperation by establishing a means of disseminating and sharing information; and recording and tracking objections and their resolutions.

The resulting general requirements for both a tracking function and a document storage and retrieval function are parallel to those of LSS. Some additional, more specific requirements were derived during the IMS conceptual design. The system should:

1) Be designed for the needs of the non-technical user,
2) Provide access to 40 simultaneous users of the tracking or structured index and 60 simultaneous users of the document collection,
3) Maximize compatibility with other related systems,
4) Be flexible in accommodating future enhancements,
5) Maintain the monitoring and tracking information consistent with the official project status and current within 30 days,
6) Provide various indexing and retrieval options including bibliographic, keyword, abstract, and full-text search of the document text,
7) Enter documents into the system within 30 days of receipt,
8) Provide a demonstrated potential for high percentage of recall (80%) and precision with recall efficiency taking precedence,
9) Provide, if possible, for ranking of documents in response to a query according to the importance to the query.

In 1987, SAIC initiated the Information Management System Bridge Program (IMSBP). The purpose of the IMSBP was to test and evaluate the technical feasibility of implementing a records management system based on full-text storage, searching and retrieval technologies (SAIC, 1988).

The Bridge Program involved time measurements for data capture, accuracy of optical character recognition processes, storage requirements for electronic images, and retrieval tests on indexed data, document text, and document images. The resulting information will prove to be valuable in the LSS conceptual design and will be referenced in later reports in this series.

2.1.5 NRC Pilot Project

In 1985 the Aerospace Corporation, under contract to the NRC Division of Waste Management, conducted a requirements analysis for a Licensing Information Management System (LIMS) which could facilitate daily operations and address the 3 year high-level waste license review process. The program led to a demonstration Pilot Project which was designed to test and evaluate the application of computer technology. The requirements analysis (Aerospace, 1986) was based upon interviews conducted with personnel in the Office of Nuclear Materials Safety and Safeguards, as well as other organizations in NRC, a review of the NRC Document Control System, and a review of the provisions in the rules for licensing proceedings (10

CFR 2). The resultant requirements were general in nature, concentrating on system functionality:

1) Comprehensive Content: Store in retrievable manner the full text of any record likely to be requested that pertains to high-level waste in compliance with 10 CFR 2. Provide capability to capture, store, and retrieve records relevant to licensing of nuclear waste during transport and for all proposed sites including permanent and monitored retrievable storage.

2) Multi-Media Accommodation: Store, index, and provide access to records on hard copy, microfiche, charts, magnetic tape, disk, and other accessible media.

3) Broad Indexing Capability: Be able to search and select by keywords, or descriptor phrases, that define the subject, author, and title and by significant words in context in abstracts and text. Be able also to search on date, issuing agency, identifying number, and other necessary identifiers. To ensure information recall and precision, the search routine must operate on both standard abstract terms (e.g., title, author, keywords) and the full text.

4) Prompt Response: Verify the existence of a record, determine the location of a record, and display, on line, the full text of records resident in the data base in real time at authorized user terminals. Produce hard copy of any record by use of high-speed laser printing. Distribute quality copies of records to users in accordance with the requirements 10 CFR 2 and 9.

5) Operational Availability: Make the LIMS available as soon as possible for the prelicensing phase and have it remain operational during licensing and after licensing.

6) Security: Protect against the loss and destruction of records and protect privileged material by controlled access.

7) Related Systems Disclosure: Disseminate general descriptions of information management systems maintained within the DOE community related to high-level waste. Provide instructions for user access to systems and provide a thesaurus of keywords and descriptors to facilitate user queries.

8) Simplicity of Use: Provide access (read and print only) for non-technical users with no prior on-line data base interaction experience.

9) Long-Term Viability: Incorporate capability to improve service throughout the licensing period.

10) Accessibility: Provide remote terminals so data bases are accessible to personnel of the States, Indian tribal organizations, and the general public.

11) Compatibility: Provide network interfacing compatibility to access other information management systems and data bases. A menu-driven interactive LIMS is required to lead the user to the on-line data bases and provide a common and unified access to the full information system.

12) Reliability: Maintain system with minimum downtime. Protect data files during system crashes and provide recovery in less than 24 hours.

13) Affordability: Provide an information management system and networking system that are cost effective.

14) Standardization: Use data transmission, library, and information science standards for data storage and records transfer for data bases as set by the American National Standards Institute.

15) Completeness: All participating parties involved with the LIMS must acknowledge (certify) that they have adopted and followed procedures ensuring that all relevant records are submitted to the LIMS.

The Pilot Project (officially termed the Transitional Licensing Support System) was implemented by NRC to begin the capture of HLW licensing-related documents and to test the feasibility of various technical concepts for information storage, primarily image capture, optical disk storage, and full text search and retrieval.

2.1.6 LSS Functional Requirements Study

Arthur Young International completed a study and issued a report on the functional requirements and design concepts for the LSS (Young, 1987) for the OCRWM Office of Geologic Repositories. The requirements identified, which are in the scope of this preliminary needs analysis (i.e., which relate to requirements other than those concerning data scope), are summarized below.

Information Access Features

1) Structured index searching. The LSS should support queries for specific types of information about documents which have been compiled through cataloging. Searches for the following types of information should be supported:

Author/originator	Originating organization
Title	Recipient
Date created	Receiving organization
Accession number	QA level
WBS number	Access restriction
Abstract	Keywords and phrases
Site applicability	Report number
Type of document	Revision number

2) Full-text searching. The LSS should support full-text searching of LSS records. Full-text searching involves finding the location of words and phrases (and their logical combination) within records.

System Performance Features

1) Search and Retrieval Features
   - o The system should provide menu-driven user access.
   - o The search criteria during a query should be able to be retained.
   - o User help should be available on-line.
   - o Bit-mapped images of documents should be available as well as ASCII text.
   - o Summary descriptions of retrieval information and counts of retrieved items should be available.
2) Document Control Features
   - o The system procedures should verify the integrity of information at the time it is captured into the system.
   - o The system procedures should verify that information has been authorized for entry at the time of capture.
   - o The system should avoid duplicate records.
   - o The information in the system should be verified for conformity with the original source document.
   - o The association between extracted cataloging information (appearing in a header, for example) and the associated document text in the system should be verified.
3) Access Control Features
   - o Access to the system should be controlled and restricted to authorized users, including members of the public.
   - o Accidental or malicious destruction and alteration of information in the system should be prevented.
   - o The system should be able to detect and prevent "unreasonable" queries which would overburden the system.
4) Acceptable Input Media
   - o The system should be able to capture information from computers and word processors in the electronic form of disks and magnetic tape.
   - o The system should also be able to capture hard copy text and figures.
5) Output Features
   - o The system should support screen output in ASCII as well as bit-map form, and should be able to highlight search information and other items of special interest.
   - o The system should be able to display both cataloged and full-text information.
   - o The system should be able to support the transmission of ASCII data for downloading.
6) Miscellaneous Capabilities
   - o The system should be able to support multi-user access from around the country.
   - o The system should provide suitable backup and recovery capabilities.
   - o The system should be operational by 1991.

2.1.7 Conclusions From Previous Examinations

An assessment of the general need for an LSS combined with the specific needs of potential users outlined in the foregoing reports leads to a number of general characteristics of a system which should be considered in the design. These characteristics include:

- A method for capturing, validating, and managing all documents pertinent to the licensing process,
- A method of ensuring the security of the system as a whole, as well as any confidential or proprietary documents, against unauthorized access and alteration,
- A retrieval mechanism that is quick, has multiple access points, and provides good recall and precision,
- A method for ensuring that all documents or document surrogates are fully searchable,
- A storage system that can provide paper, microform, or electronic media copies of any document in the system,
- A method for rapid access to the information stored, support of simultaneous usage, and ease of use for both the novice and experienced user,
- A provision for the conversion to future technologies, while maintaining compatibility with systems currently in use.

These characteristics support the needs expressed by NRC for meeting the three-year licensing timetable. It is evident that while these conclusions provide a firm conceptual basis, they do not present any quantifiable detail on which to base system design specifications. Further, these examinations were undertaken before the NRAC began its proceedings and, therefore, could not consider the imminent Rule. Neither have other non-end user requirements on the LSS been considered in great depth to date. The following two sections address these requirements.

2.2 Negotiated Rulemaking Advisory Committee Requirements

The Federal Register notices (Federal Register 1986, 1987) establishing the intent to form a HLW Licensing Support System Advisory Committee (generally referred to as the Negotiated Rulemaking Advisory Committee or NRAC) for negotiated rulemaking contained some information pertaining to the requirements for an LSS which form a basis for their discussions. These requirements as envisioned by the NRC can be summarized as follows:

- Capture in electronic form all of the data that would normally be generated to support the licensing decision.

- Contain a "no-access" file for privileged data with appropriate safeguards.

- Provide open access to all parties, with the exception of data in the privileged file, available at minimal cost to the user.

- Facilitate review of the information through the provision, to the extent practicable, of full text search capability.

- Provide for electronic transmission for submission of motions and other documents associated with the licensing proceeding.

As noted in the Introduction, representatives of all parties began meeting in September, 1987 to negotiate a rule that would determine the requirements of the LSS and thereby expedite the necessary revisions to 10 CFR 2 relating to Rules of Practice for Adjudicatory Proceedings.

The NRAC met monthly through December, 1987, when Congress passed the Nuclear Waste Policy Amendments Act and substantially altered the scope of the nuclear waste management program. As a result, the membership of the Committee decreased substantially and the January and February, 1988 meetings were canceled, pending a reorganization of the Committee's agenda.

The Committee currently consists of representatives from the following organizations:

- State of Nevada
- Department of Energy
- Nuclear Regulatory Commission
- Edison Electric Institute and Utility Nuclear Waste Management Group
- Coalition of non-profit environmental groups (Sierra Club, Environmental Defense Fund, Friends of the Earth)
- Local governments from the vicinity of the Yucca Mountain site.

Other organizations may petition for membership at any time. The meetings are open to the public and public comment is invited.

2.2.1 Negotiated Rulemaking Process

The process which has been used by the Committee is based on achieving a consensus, which NRC is then bound to use in development of the final rule. The steps in the development of consensus include:

1) Establishment of the procedures for the process

2) Education of the members in the licensing process and information management

3) Identification of the issues

4) Negotiation of the issues

5) Draft of the rule

6) Consensus on the rule.

The September meeting was primarily procedural (Step 1), and the October meeting focused on Steps 2) and 3). The meetings in November and December consisted of negotiation of some issues with the appearance of tentative consensus being reached. NRC has elected to expedite the negotiation process by submitting their position on several issues to the Committee, which has become the reference point for discussion. The issues which have emerged can be summarized as follows:

1) What documents must be in the LSS? (i.e. discoverable)

2) What subset of these documents must be in full-text and when must this be completed?

3) How will privileged material be handled?

4) How will drafts, handwritten material, and marginalia be handled?

5) What are the mechanics and responsibilities for record entry?

6) What are the procedures governing access to LSS - who, how, and at what cost?

7) What procedures are required for dispute resolution prior to the license application submission?

8) Who will have administration and oversight over LSS?

To date, Items 1, 2, and 3 have been discussed in detail with a tentative consensus emerging on these issues. The remaining items have been addressed only briefly by the group as a whole.

The remaining schedule for the process is as follows:

March - negotiation and "tentative" consensus of all issues

April - draft language of the rule

May - agreement on language of rule

June - final consensus on rule

2.2.2 Identification of Needs

Based on OCRWM's and contractors' observation of the negotiated rulemaking process and the discussions to date, it is our opinion that certain requirements and needs are emerging from this process. As noted in the schedule above, the March meeting will provide significant further insight into the consensus position. While the final position of the

Committee could be different than the current status would indicate (which is the reason for the term "tentative" consensus on the issues), we believe the following needs can be identified pending final consensus:

- Different records within the LSS may require different treatment in cataloging and indexing. The LSS should include the capability for headers including subject terms or keywords and abstracts, and full-text search although not all records will use all of these indexing methods.

- Probably most, but not all documents will have the text entered into LSS and indexed for search.

- The filing of motions and other documents associated with the hearing will be facilitated by an electronic mail capability.

2.3 <u>Other Non-User Requirements</u>

2.3.1 Standards and Quality Assurance

Generally speaking, all records that are part of LSS must be legible, identifiable, and retrievable. They must be stored in a manner providing protection against damage, deterioration, or loss, and requirements must be set for access, retention, maintenance, and disposition. Certain of these requirements have been mandated by OCRWM, National Archives Records Administration (NARA), General Services Administration (GSA), National Fire Protection Association (NFPA), and the American National Standards Institute (ANSI) in conjunction with the American Society of Mechanical Engineers (ASME). They appear in such documents as the Code of Federal Regulations, NARA and GSA Bulletins, OCRWM documents, and NFPA and ANSI/ASME Standards, some of which are summarized below. These requirements place additional non-user requirements on the LSS.

Since the LSS is to be the repository for OCRWM records, it must be operated in accordance with the quality assurance (QA) plans that would normally be followed by that Office. Therefore, any activity affecting the quality of the records and their computer surrogates must be controlled through the use of written procedures (Quality Assurance Plan). In addition, records must be kept to provide evidence of these quality-related activities. Quality assurance procedures, responsibilities, and criteria are detailed in OGR's Quality Assurance Plan for High-Level Radioactive Waste Repositories; ANSI/ASME NQA-1 (1983), Quality Assurance Program Requirements for Nuclear Facilities; and 10 CFR 50, Appendix B, Quality Assurance Criteria for Nuclear Power Plants and Fuel Reprocessing Plants. Some responsibilities include performing periodic QA audits; controlling the removal of records from the system; controlling access to the records; protecting records against damage, deterioration, or loss; and developing indexing for prompt retrieval of documents. Procedures to be developed include those for record validation, record storage access and protection, filing and indexing of records, and removal of documents from storage. Minimum criteria for these procedures and responsibilities are specified in these documents. (DOE, 1986)

A Records Management Plan must also be prepared which will identify the types of records to be entered into the system, methods which will control in-process records, and methods and schedule for purging non-permanent records. All QA records to be entered into LSS must be identified by a unique number or other similar designation. In addition, QA final reports are to contain a listing, by unique record number, of all documents used to compile or evaluate the report. This listing is to include all referenced documents, as well as review documents, computer codes, data sheets, etc. All of these referenced documents shall be retrievable from the system, with the exception of readily available publications such as encyclopedias, dictionaries, handbooks, etc. (DOE, 1986)

Records for storage in the system may be provided by any system user. However, only designated agents are authorized to enter data directly into the system. In addition, measures are to be established to assure that only these authorized personnel have access to the computer and archives. The system must include features to prevent unauthorized access and willful or accidental damage to the data base contents or archival materials. (DOE, 1986)

A determination must be made at the time a document is received as to which form of the document (paper, microform, or electronic) will be the "record" of that document. All records, whether paper, microform, or electronic, can be destroyed only according to an approved retention schedule (NARA, 1987). Since most of the records in this system are post-closure (retention of 300 to 1000 years) or lifetime records (retention for the operating life of the repository), it is necessary to determine which form of the document is to be the "record" to ensure against inadvertent loss (DOE, 1986). All records are to be controlled from the time they are received until they are stored in a permanent storage facility or, in the case of non-permanent records, disposed of in the designated manner.

All record storage media, whether electronic, microform, or magnetic, must be properly labeled. The minimum amount of information on each label for electronic and magnetic media includes title, dates, software and file code, and identification of the equipment on which the records were created (36 CFR 1234). Microform is to be labeled in a similar manner (36 CFR 1230). All records are to be indexed to assist in locating the records; indexes should be determined by those characteristics which would assist in distinguishing one record from another. Records should be easily retrieved by authorized personnel throughout the retention period.

Records that are documentation of experiments or research shall be contained in bound logbooks or other suitable means. Entries in these documents shall consist of those items described in Supplement No. 5 to the OCRWM Quality Assurance Plan for High-Level Waste Requirements, items such as description of the experiment's objective, qualifications of the participating individuals, identification of equipment and materials used.

Both NARA and GSA have set guidelines and requirements for the storage and handling of magnetic media; these are detailed in 36 CFR 1234. In general, magnetic media are to be tested before use and annually sampled for loss of data. Error rates are not to exceed ten before the data is transferred to a new tape and data restored if possible. Specific documentation is to be supplied with each diskpack or reel of tape. Records on disks or diskettes are to be transferred to magnetic tape for permanent storage. Duplicate copies of media are to be stored off-site. Data are to be stored on National Bureau of Standards (NBS) mandated media.

Similar standards for microform records have been detailed in 36 CFR 1230. Before replacing original records by microform, steps must be taken to ensure that the microform is an adequate substitute for the original. The microform must be of archival quality. Inspection of microforms are to be scheduled every two years using a randomly-selected sample. Microform that is found to be deteriorating shall be replaced. Specific instructions for disposing of microform are given in 36 CFR 1230.26.

Standards for physical storage facilities for either temporary or permanent storage of archival materials are included in 36 CFR 1228 Subpart J. These requirements are set to guard against fire, theft, and deterioration. Since these requirements do not directly affect the records themselves, they will not be discussed in this report.

2.3.2 Schedule

Two driving forces determine the time frame in which the LSS should be operational. The first is the contractual requirement between DOE and development contractor (SAIC) to provide for a system operational by August of 1990. (The schedule has been modified from the original RFP (DOE, 1987) to accommodate the current requirements analysis). The definition of operation consistent with the August 1990 date is to demonstrate "successful loading and system operation of up to 4,000,000 pages of real data." The date upon which the LSS will be fully operational (i.e. contain all backlog data and maintain loading of current data) will be a function of the backlog volume and the rate at which it can be absorbed.

The second driving force is the programmatic requirements to use the LSS prior to the license application submittal to NRC. The primary users of the LSS at this stage would be DOE personnel preparing the licensing documents and potential parties to the hearing who would participate in the discovery process. Given that the contractual data of August 1990 is over 4 years prior to the anticipated application submittal (see Section 4.2.2), it would appear that these dates are consistent and the LSS development schedule is in accordance with program requirements.

2.4 Direct Needs Evaluation from Potential Users

Several of the studies summarized in Section 2.1 relied on interviews to collect data on LSS requirements. These interviews, however, were generally intended to compile licensing process and program understanding, and not to determine the individual needs of representative potential LSS users. The studies could not therefore address user needs in sufficient detail to provide the basis for LSS system specifications. The analysis presented here has attempted to refine that understanding of needs by direct contact with potential users, a process which has also been enhanced over previous attempts by the fact that potential users have become more educated in the facets of on-line retrieval systems. The following two sections of this report discuss the results of these interviews (Section 3) and estimate their impact on geographic and temporal demand on LSS (Section 4).

3.0 SURVEY OF POTENTIAL USER NEEDS

3.1 Identification of Usage Groups

In order to make valid assumptions about the use of LSS, the users were divided into categories: engineers, lawyers, managers, intervenors, etc. Recognizing that a single user may use the LSS in more than one way (i.e., a lawyer might be seeking technical as well as legal material), the users were grouped into four usage pattern categories, reflecting similar traits:

Technical and Engineering Usage
Regulatory and Licensing Usage
Management and Administrative Usage
Public Information and General Public Usage

It also became apparent that, in addition to these end-users of LSS, there were two other usage groups:

Intermediary Usage
Data Base Management and Quality Assurance Usage

Although the intermediaries would be searching for the same information as the end-users, the way they used the system to search and retrieve the information and the amount of time spent on the system for each search is expected to be very different. Similarly, the way in which the Data Base Management Group uses the system would be different from all the other groups. A description of the usage category, examples of who would fall into the category, the type of information sought, the subsystems of interest (secondary interests are in parenthesis), and the query approach of these users was developed for each usage pattern. A questionnaire that could be used for all groups (except the Data Base Management Group) was then developed. Since the demands on the system of this latter group would be radically different from the other five, a different questionnaire was used. The description of each of the usage categories follows.

Name of Usage Category: TECHNICAL AND ENGINEERING USAGE

Generic Description: Usage in this category is expected to be primarily by scientists and engineers requiring information during the preparation or review of technical reports used in support of the licensing process. This group will be mainly the technical staff of federal agencies, national laboratories, state and local agencies and environmental and public interest groups. Their questions will deal mostly with primary data, published analyses of technical issues, computer program documentation, QA procedures and testing procedures.

Query Approach: The thinking of this group is analytical, experimental, and scientific. They are generally concerned about everything written on a specific narrow topic or are looking for a specific piece of data to support a hypothesis, experiment, or test. Their interest is also on who authored the information or on what other information was published by the same author.

Examples: Members of NRC's Federally Funded Research and Development Center (FFRDC) technical staff developing independent verification of site performance claims appearing in the repository construction authorization request.

DOE national laboratory technical staff under contract to develop site characterization information needed to show compliance of the site with 10 CFR 60.

Technical and scientific consultants hired by public interest organizations to independently verify DOE site and repository performance claims.

Information Sought: Some of the documents sought in this usage category are DOE, NRC and national laboratory technical reports, articles in scientific and engineering journals, progress and summary reports of contract for government agencies.

LSS Subsystems: Records Access Subsystem, (Regulations Access Subsystem, Issues and Commitments Tracking Subsystem).

Name of Usage Category: REGULATORY AND LICENSING SUPPORT

Generic Description: Usage in this category is expected to be primarily by regulatory and licensing specialists (including legal staff) requiring access to both technical and regulatory information. Before submittal of the license application, this group will perform three major regulatory functions. First, regulatory support staff will perform an ongoing oversight role to ensure that technical work will result in a complete and defensible license application. Second, the regulatory support staff will direct and participate in topical report development, seeking early resolution of issues. Third, programmatic decisions must be reviewed by legislative/policy analysts to determine if actions contemplated are within the letter and intent of federal, state, and local laws and regulations.

After submittal of the license application, the licensing support staff will be responsible for developing positions on licensing issues, identifying witnesses, preparing testimony, responding to motions, etc. One important aspect of the hearing is the conduct of discovery, a process which allows all parties equal access to relevant information.

Query Approach: These users are procedure- and strategy-oriented, with a broad qualitative bent. Their concerns are with defensability of positions, completeness of documentation, and direction of overall policies and strategies.

Examples: Members of the DOE regulatory staff developing topical reports on regulated technical issues such as those presented in 10 CFR 60.

Intervenors requesting all information regarding the basis for design decisions.

NRC licensing staff assessing the completeness of license application documents.

Information Sought: Some of the documents sought in this usage category are technical reports, correspondence containing review comments, technical meeting minutes, regulations, regulatory guidance, planning documents, and commitments.

LSS Subsystems: Regulations Access Subsystem, Records Access Subsystem (Issues and Commitment Tracking Subsystems).

Name of Category: MANAGEMENT AND ADMINISTRATIVE USAGE GROUP

Generic Description: Usage in this category is expected to be primarily by managers and administrators who are concerned with projects and contracts they are conducting or monitoring. This group will be mainly task, project and program managers and administrators as well as line managers, at government agencies, national laboratories and private contractors. Their questions will deal mostly with project information such as commitments, obligations, deliverables, schedules and progress reports. (Note that no financial information is to be stored in the LSS.)

Query Approach: This group is concerned with meeting schedules and commitments, managing projects and meeting deadlines for milestones, and monitoring contracts. They will generally be looking for specific information and specific documents.

Examples: NRC contract administrator responsible for administration of technical contract with private sector firm for supporting independent development of performance assessment capabilities at NRC.

Administrative assistant to OCRWM branch chief responsible for developing a portion of repository construction authorization request to NRC.

Staff member of private contracting firm (working for NRC or DOE) responding to request for information from firm's management, concerning contract milestone schedules.

Information Sought: Some of the documents sought in this usage category are planning documents, cost and schedule performance charts, statements of work (SOWs), QA audit reports, correspondence, action and commitment tracking documents, and memoranda of understanding.

LSS Subsystems: Issues and Commitments Tracking Subsystem, Records Access Subsystem, (Regulations Access Subsystem)

Name of Usage Category: PUBLIC INFORMATION AND GENERAL PUBLIC USAGE

Generic Description: Usage in this category is expected to be primarily in support of information needs of the general public, either in response to direct inquiry or through dissemination by public information specialists. This group will be mainly public information officers, public document room staff, members of educational and other public institutions, reporters, community members, civic activists and members of concerned citizens groups. Their questions will deal mostly with general and descriptive information about nuclear waste management and OCRWM activities, and summary information on technical and environmental issues.

Query Approach: These users are thinking in broad, general terms, their information requirements usually focused on descriptive and issue-related material for personal or local applications or to disseminate to a wide range of constituents.

Examples: OCRWM Public Information Staff, responsible for the preparation of information brochures, briefings, etc.

Public information and (non-technical) research staff of the Sierra Club, researching background of points developed in the repository construction permit request.

Members of affected communities and Indian tribes, researching questions of specific local or general background interest.

Information Sought: Some of the documents sought in this usage category are records of public hearings, issue papers, fact sheets, documents open for public comment, and press releases.

LSS Subsystems: Records Access Subsystem, (Regulations Access Subsystem)

Name of Usage Category: INTERMEDIARY USAGE

Generic Description: Usage in this category is expected to be primarily in support of information needs of all usage categories, generally in response to inquiries by those who do not have direct access to LSS or those who do not want to access the system themselves. This group will be mainly librarians and information specialists and may also include administrative assistants, researchers and paralegals who have a working knowledge of information retrieval or have become thoroughly experienced in searching the LSS.

Query Approach: These are service-oriented people who are patient, persistent, and curious enough to try many methods of obtaining information for others. They generally have a broad understanding of the organization of information to assist them in their task. Their concerns are more with the methods of extracting information from the system than with the content.

Examples: Information specialist in OCRWM Public Information Office

Reference librarians at DOE libraries in Germantown, Oak Ridge, Washington, D.C. and at DOE national laboratories.

Paralegal staff at an intervenor law firm.

Information Sought: The documents sought in this usage category will be all of the documents in the system needed by each of the four end-usage categories.

LSS Subsystems: Records Access Subsystem, Regulations Access Subsystem, Issues and Commitments Tracking Subsystem

Name of Usage Category: DATA BASE MANAGEMENT AND QUALITY ASSURANCE USAGE

Generic Description: Usage in this category is expected to be primarily one of controlling and facilitating the flow and quality of data and documents into and out of the LSS. This group will be mainly QA/QC staff, data base maintenance staff and data base-use trainers. Their questions will deal with support activities such as quality control audits, error corrections, data base maintenance and maintaining an up-to-date thesaurus.

Query Approach: This groups is concerned with the operation of the system--how to get information in and out, what checks and balances to perform, etc. They are generally systematic, detail-oriented, and persistent in tracking errors and problems.

Examples: A data base administrative staff member verifying that the system contains all of the referenced attachments to a correspondence previously entered in the system.

A data base administrative staff member verifying that a document meets all system acceptance criteria (e.g., legibility and completeness) before being entered into the LSS.

A data base administrator performing internal consistency and QA checks on the content of document header information for data base maintenance.

Information Sought: The documents sought in this usage category will be all of the documents in the system needed by each of the four end-usage categories.

LSS Subsystems: Records Access Subsystem, Regulations Access Subsystem, Issues and Commitments Tracking Subsystem.

3.2 LSS Needs Evaluated

As indicated in Section 1.1, the scope of this analysis is limited to requirements not related to the amount or type of information stored in the LSS. These requirements can be classified as pertaining 1) to system capabilities and performance requirements or 2) to information access features of the system. The following subsections list the specific topics within these two categories for which data were collected during the interviews. Questions were not asked directly on most of these topics, so the respondent would not be prompted to "need" whatever was possible. Rather, the questions were directed toward the requirements of the job performed by the respondent.

3.2.1 System Performance Requirements

Information was sought on the system performance topics listed below. The data on these system performance requirements are needed for communications and computing capacity sizing, communications topology, terminal distribution, output capacity and distribution, security design, QA procedures development, and selection of operational features.

I) User session characteristics
  1) Average session characteristics
    a) session length
    b) sessions per day (or per week)
      - per user
      - per installation (site)
  2) Peak session characteristics
    a) session length
    b) sessions per day
      - per user
      - per installation (site)
    c) when is peak expected?
      - time of day
      - during licensing processes

II) User geographic distribution
  1) Number of users
    a) total number of users at site
      - initially
      - peak
    b) number of simultaneous users
      - initially
      - peak
    c) what organization does this represent
  2) User location
    a) city
    b) site / building
    c) are limited capabilities acceptable at remote locations?
    d) maximum acceptable waiting time to wait to get on a terminal
      - routinely
      - priority / urgent

III) Response time characteristics
- 1) Interactive session (maximum and routine acceptable delay until system begins to respond)
  - a) during large indexed search
  - b) during large full-text search
  - c) while paging a document
  - d) in a Tracking Subsystem
- 2) Hardcopy (maximum and routine acceptable time to receive, for print jobs >100 pages and <100 pages)
  - a) copies of documents (or parts of documents)
    - time to receive <= 100 pages
    - time to receive > 100 pages
  - b) material other than documents (header data, etc.)
    - time to receive <= 100 pages
    - time to receive > 100 pages

IV) Output capability needs
- 1) To terminal
  - a) amount of document text per screen
  - b) document images needed?
- 2) Hardcopy
  - a) need for print capability at terminal (or is site enough)
  - b) quality needed (dot matrix or laser)
- 3) Downloading
  - a) peak file size
  - b) average file size
  - c) interactive or batch (overnight)
- 4) Other forms of output
  - a) diskettes
  - b) magnetic tape
  - c) microform
  - d) other forms

V) Other features
- 1) Need for electronic mail on system
- 2) Need to be able to annotate text on system

3.2.2 Information Access Features

Information was sought on the information access topics listed below. The data on these system performance requirements is needed for the selection of retrieval software features, data base architecture, header and cataloging procedure design, and training planning.

I) Which access technique would user need most?
- 1) Full-text search
- 2) Structured index searching
- 3) Combination
  - a) both always available, or
  - b) availability limited by size of search
    - tradeoff point (size or number of "hits")
    - speed limitations, if any

II) Full-text searching features
  1) Need to full-text search
    a) maximum amount of text at one time
    b) nature of information to be searched
      - document text only
      - document text and all header information
      - document text and abstract only
  2) Full-text features needed
    a) Boolean logic
    b) proximity relations
    c) misspelling tolerance

III) Structured index searching features
  1) Data base entry points (header fields needed)
  2) Which need controlled vocabularies

IV) System/end-user interface
  1) Direct access
    a) simple prompted
    b) prompted and command language
    c) expert system assisted
  2) Access via trained operator

3.3 Description of the Interview Sessions

During the period from 25 January to 2 February 1988, SAIC staff conducted 44 interviews with potential LSS users and with people having significant insight into the behavior of future users. Most of the people interviewed are affiliated with the NRC, DOE, and NRAC; some surrogate users were also included. A complete list of these people and their affiliations is included in Appendix A.

The majority of interviews were conducted face-to-face; the others were conducted by telephone. The sessions averaged over an hour each. Interviewers used a sequenced check list of topics and questions but were given considerable discretion in allowing the people interviewed to have their full say and to change topics at will. The purpose of this interview style was to put respondents at ease and encourage them to be frank.

A variety of both facts and opinions were collected during the interviews and will provide a continuing source of assistance in planning for the LSS. The immediate concern is with the rich statement of needs acquired by the interview process. Some of the information is quantitative and lends itself to concise presentation. Other information is less easy to quantify. The latter information is presented here only when it reflects a position taken by several respondents and represents a trend of opinion expressing a specific need. Many respondents asked to be involved in other LSS data collection and planning activities. Contact with these people will be maintained and they will be involved to the greatest degree possible.

3.4 Results of Interviews

The interviews usually covered five main topics:

- Expected LSS User Session Characteristics: the average and peak number of sessions per day, typical session lengths, and the anticipated peak both in terms of time of day and events in the licensing process.

- Distribution of Users: the total number of users at a site and the number of simultaneous users at a particular site, both during initial use of the system and at peak times.

- Response Time Requirements: the time it would take the system to respond while conducting a search, as well as the time it would take to receive hardcopy of a desired document, either whole or in part.

- System Output Requirements: output to terminal screen, to printer, or to download to other systems; other forms of output including diskette, magnetic tape, microform, electronic mail, or other user-defined output.

- LSS Information Access Needs: available access (full-text, structured index, or both), searchable information (document text, abstract, headers), search mechanisms (Boolean logic, proximity relationships, misspelling tolerances), and data base entry points (author, document type, technical discipline); system/end-user interfaces (menus, expert system assists, intermediaries) were also included.

Table 1 shows some of the interview findings that can be conveniently quantified. This table combines and summarizes the detailed results obtained for the six usage groups used in sequencing the topics described in the following paragraphs. The detailed results and further refinements obtained from prototype tests and other sources (see Section 5.0), will constitute the requirements basis for the LSS system specifications.

TABLE 1: Summary of Some Quantitative LSS Performance Requirements as Seen by Potential Users

System Characteristics	Minimum Value	Maximum Value	Central Tendency of Response
<u>User Session Characteristics</u>			
Average Session Length	5 mins.	All Day	65 mins.
Peak Session Length	5 mins.	All Day	90 mins.
Avg. No. of Sessions	None	Continuous	1/day
Peak No. of Sessions	4/year	20/day	3/day
Peak Time of Day	Early AM	Late PM	Mid-Late AM
When Peak Occurs	--	--	Major Licensing Events
Max. Acceptable Wait			
to log-on - Routinely	2 secs.	2 days	2 hrs.
to log-on - Priority	2 secs.	2 hrs.	1/2 hr.
<u>Acceptable Interactive Response Times</u>			
Large Indexed Search			
Max. Delay at Peak Demand	1 sec.	30 mins.	12 mins.
Acceptable Routine Delay	1 sec.	24 hrs.	1/2 hr.
Large Full-Text Search			
Max Delay at Peak Demand	15 secs.	30 mins.	15 mins.
Acceptable Routine Delay	2 mins.	1 day	1/2 hr.
While Paging a Document			
Max Delay at Peak Demand	1 sec.	1 min.	10 secs.
Acceptable Routine Delay	1 sec.	1 hr.	15 mins.
<u>Hardcopy Receipt Times</u>			
Documents			
100 pages or less	5 mins.	2 days	1/2 day
More than 100 pages	4 hrs.	3 days	1 day
Header and Other Data			
100 pages or less	5 mins.	2 days	6 hrs.
More than 100 pages	5 mins.	2 days	1 day

3.4.1 User Session Characteristics

The average potential user anticipates using the LSS computer systems about once each day for an average session length of about an hour. The variation among potential users is great, ranging from 5 minute sessions to those who expect to be on the system the entire day. The Technical and Engineering and the Regulatory and Licensing Support Usage Groups represent most of the potential users. While average total use would be similar between these groups, interview data shows that the Regulatory and Licensing Support Group would have session needs of substantially longer duration than the Technical and Engineering Group, but would need only about half the number of sessions as their technical counterparts. Usage by all groups shows that there would be a marked increase at peak times during the licensing process, with a likelihood of the number of sessions tripling and session length increasing to about an hour and a half. These peaks would occur daily from mid- to late-morning and from mid- to late-afternoon and are expected to be intense when the preparation, issuing, and review of key licensing documents are being conducted. (See Section 4.2 for a more detailed discussion of these results.) Under routine circumstances, potential users would be willing to wait almost two hours to have a terminal available, but under priority conditions, they would want access in about 25 minutes.

3.4.2 Acceptable Interactive Response Times

The tolerance for delays in system response reflect the kinds of searches being conducted by users. Simple prompts or requests (such as paging through a document or searching a tracking system) would be expected to be processed quickly (an average of 20 seconds and 15 minutes, respectively) while longer searches using a large index or full-text would be expected to require more time (about a half hour each). Longer delays under peak demand appear tolerable to potential users. On the average, they would be willing to wait about twice as long under these conditions. Representatives of the Management/Administrative usage group are much less patient than those of the other groups. They also insist, as do many of the other users, that the system provide some feedback quickly and allow the user to go on to other work while the search is being conducted. A desire was frequently expressed for some means of feedback as to the expected delay associated with a given request.

3.4.3 Hardcopy Output

The requirements for receipt of hardcopy as an output of LSS varies very widely. Managers would be willing to wait 3 hours on the average for documents of 100 pages or less. Technical users would wait 7 hours and regulatory staff over 22 hours for the same materials. For hardcopy more than 100 pages most users would wait overnight but regulatory staff would be content with about a two-day delay. The tolerable waiting time for hardcopy of header and tracking information is, across all groups, about half the time of the wait for small documents.

3.4.4 Print Requirements

With very few exceptions, potential users would require a print capability. The typical need is for a dot matrix printer at or near the terminal. Laser printers would generally be required on the same floor or in the same building. Several technical users stressed the importance of laser printers to produce required diagrams and maps.

3.4.5 Downloading

About half of all respondents stated requirements for downloading LSS information to other systems. Average and peak downloads generally ranged from 2-3 pages to 20-30 pages. However, some regulatory staff indicated peak load requirements in excess of 1,000 pages.

3.4.6 Other Forms of Outputs

Only a few of the respondents specified other than hardcopy output. Diskettes were the most frequently mentioned but fewer than 25% of respondents so indicated.

3.4.7 Other System Features

More than 90 percent of potential users said that they would consider electronic mail to be a useful LSS service. The ability to annotate text on the system was favorably evaluated by about half of the respondents. However, many, both pro and con, were very concerned that, if implemented, such a capability would be able to assure privacy and security. The tone and frequency of this concern was much noted by interviewers.

3.4.8 Information Access Feature

Potential users were asked about their full-text searching and structured-index searching needs. More than 90 percent of the respondents indicated a preference for having a combination of both available. Of this large majority, about half felt that the combination search capability should always be available while the other half believed that the availability of full text searching could be limited by size of search. Boolean constraint and proximity-type features were consistently supported. Misspelling tolerance features were heavily favored, but most of these respondents insisted that an on-off switch for this feature be provided.

3.4.9 Data Base Entry Points

A long array of header fields were provided to respondents who were asked to indicate which fields they thought were important to their expected uses. A header is defined as cataloging information appended to the beginning (or "head") of a document and can consist of such fields as author, corporate affiliation, journal name, number of pages, etc. The following field descriptions were found desirable by a wide majority:

- Technical Discipline and Subdiscipline
- Originating and Recipient Organization
- Dates the Document Was Created
- Author
- Issue (from Hierarchy)
- Document Type
- Baseline Data Flag
- Cross-references for NRC and DOE Document Numbers.

Mixed support was provided for:

- Geographic Reference
- Technical Level
- Commitment Status
- WBS Number.

Several respondents indicated the usefulness of the NRC Division of Waste Management Technical File Plan Index. Others made reference to adding an indication of the level of quality assurance.

Respondents were also asked to indicate which fields should be subjected to controlled vocabularies. Those fields judged to be most valuable as entry points were indicated as requiring controlled vocabularies. In addition, many respondents emphasized that controlled vocabularies can be very helpful.

3.4.10 System/End-User Interface

Most respondents favored accessing the system through a prompted and command line user interface, over a simple-prompt capability. Almost all respondents suggested there would be need for some kind of expert-assisted interface, but most rejected the idea of having access restricted to only trained operators.

3.4.11 Intermediary, Public Information and Data Base Maintenance Usage Groups

Our sample of Intermediaries were mostly librarians. All Intermediaries were very familiar with searching computerized data bases, including full-text, bibliographic citations, and citation/abstract combinations. All of these potential users stressed the need for user-friendly interfaces, provided a great deal of helpful details on pit-falls in searching, and consistently emphasized the need for controlled vocabulary and the on-going training of users. Intermediaries would prefer a system where menus or expert-assisted features could be bypassed by experienced users.

The Public Information Support group emphasized the importance of tracking issues and the need of prompt entry of documents into the data base. While abstracts were not much supported by other users, this group and the Intermediary group stressed their usefulness, especially in narrowing searches down to a manageable number of hits before looking at full-text.

Data base maintenance users were concerned with procedures for loading the data base, changing and adding fields, controlling access to validated users, and providing QA activities to ensure the integrity of the data base and computer system. Much of their commentary focused on the good and bad features of their own systems. This experience will be useful in later LSS design activities.

3.4.12 Number of Users

At the present time, several years before peak activity in the repository licensing process, estimates of the number of LSS users must have appreciable uncertainty. The relative proportion of usage in the categories identified above is somewhat more certain. The following estimate of this distribution is based on discussions with key respondents during the interviews:

Technical and Engineering Usage	45%
Regulatory and Licensing Usage	25%
Management and Administrative Usage	5%
Public Information and General Public Usage	5%
Intermediary Usage	18%
Data Base Management and Quality Assurance Usage	2%

Although extremely uncertain, the magnitude of the number of LSS users must be estimated so that preliminary system sizing estimates can be made. These estimates are similarly based on interview discussions, but also consider DOE and NRC staffing plans and budget projections for organizations involved in licensing. The following estimates have been made for the number of users in two categories at peak loading during the licensing processes:

DOE HQ and Project Staff and DOE Contractor Staff	200 $\pm$ 75 users
NRC HQ, NRC Field and FFRDC Staff and State, Indian Tribe and Intervenor Usage	150 $\pm$ 50 users

Some of this uncertainty results from the ambiguous definition of "user". If occasional users are considered (i.e., everyone who would be issued a user name and account number), the number of users is probably near the top of this range. If only moderately heavy users are considered, the value would be closer to the estimate. It is important to continually resolve this value (and its geographic distribution). Note that, although the recent functional requirements report (Young, 1987) did not directly estimate the number of LSS users, the estimate of 188 terminals needed to support Washington, D.C. and the Nevada site only is consistent with the estimate above if a user to terminal ratio of about 2:1 is used.

4.0 DISTRIBUTION OF NEEDS

As pointed out by several of the studies summarized in Section 2, both the broad geographic distribution of users and the long time period the system must function are significant challenges in the design and implementation of the LSS. This section presents estimates of the geographic and temporal distribution of the need to access LSS.

4.1 Geographic Distribution

Information on the geographic distribution of usage demand is necessary for determining the distribution of LSS terminals and the design of the supporting communications network topology.

The DOE, the NRC and their contractors are expected to comprise the vast majority of LSS users. The main concentrations of these users are expected to be in:

	DOE	NRC
Washington, DC Area	Headquarters M&O Contractor HQ Technical Services Contractor Other contractors	Headquarters FFRDC Contractors Misc. contractors
Las Vegas, NV	NNWSI Management Support Contractor NTS contractor Other contractors	---
San Antonio, TX	---	FFRDC contractor

Figure 1 shows these three locations, as well as ten other locations across the country which were identified during the interviews (discussed in the previous section), as potential locations for LSS users. These interviews suggest that in the 1990 - 1992 time frame, the distribution of usage among these locations is approximately:

Washington, DC Area	50%
Las Vegas, NV	30%
San Antonio, TX	10%
Other Locations	10%

These values should be considered only estimates, since several assumptions on program and contractor stability have been made. One such assumption is

Figure 1. Geographic distribution of anticipated LSS usage. Principal usage is expected to be centered in the Washington, D.C. area (50%), Las Vegas, NV (30%), and San Antonio, TX (10%).

that the LSS usage by NRC FFRDC staff would be evenly divided between San Antonio and the Washington, D.C. area.

4.2 Needs as a Function of Time

The information on the temporal distribution of usage demand, coupled with information on the number of users (per usage location), is needed in sizing system computing capacity, sizing the width of links in the system communication network, and determining the number of terminals needed per location. Because of the extremely long period during which the LSS must function, estimates of long term demand variations is also important in designing the system. This section presents estimates of the average daily demand on the system and of the variations on this demand as the repository licensing progresses.

4.2.1 Access Needs as a Function of Time of Day

The information collected in the interviews discussed in Section 3 was used to estimate the distribution of usage demand during an average day in the 1990 - 1992 time frame. This distribution, presented in Figure 2, is based on responses to the following:

- o Average session length
- o Average number of sessions per day
- o Preferred time of day for working on the system.

This information on daily demand was combined with the estimated geographic distribution of usage given in Section 4.1 (to adjust demand to a single time zone). Each person interviewed was assigned to represent one or more usage group/location combination in the following correlation matrix. This matrix associates the geographic distribution estimates from Section 4.1 with the usage group distribution estimates from Section 3.4.12. Note that no weight was assigned to Data Base Management Usage Group users (shown in parentheses in the table) in San Antonio and at Other Locations, since no LSS computer center is anticipated at these locations. Data Base Management users were assumed to be distributed between Washington, D.C. and Las Vegas in a 3:1 ratio. The results in Figure 2 are not sensitive to these assumptions.

		End Usage Groups			Support Usage Groups		
	All Groups	Tech. 45.0%	Reg. 25.0%	Manag. 5.0%	Pub. 5.0%	Inter. 18.0%	D.B. 2.0%
Location:							
Washington, DC	50.0%	22.5%	12.5%	2.5%	2.5%	9.0%	(1.5%)
Las Vegas	30.0%	13.5%	7.5%	1.5%	1.5%	5.4%	(0.5%)
San Antonio	10.0%	4.5%	2.5%	0.5%	0.5%	1.8%	(0.0%)
Other Locations	10.0%	4.5%	2.5%	0.5%	0.5%	1.8%	(0.0%)

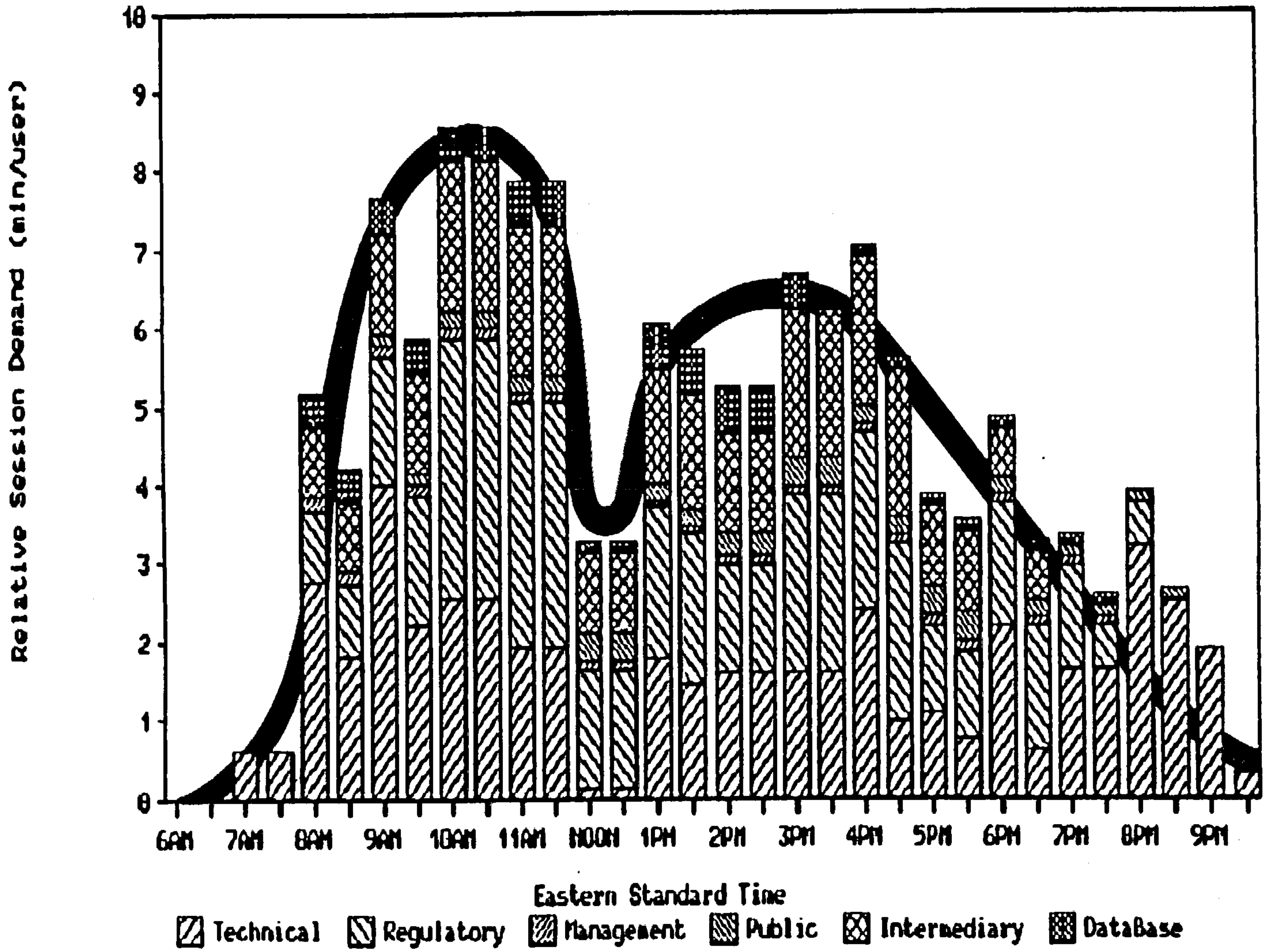


Figure 2. Routine daily usage demand histogram (in minutes per user) adjusted to Eastern Standard [illegible]e.

The average daily demand profile of each usage group/location combination was determined by averaging the profiles of the representatives sampled (so that sample size does not influence results). This was then weighted by the factors in the matrix above and summed for each usage group. The cumulative histogram of these results is given in Figure 2. The y-axis in the histogram is in minutes of demand per user for each half-hour interval. These results, multiplied by the total number of LSS users (at a given time, estimated in Section 4.2.2) gives an estimate of total demand for the system, in minutes per half-hour during the day.

It must be emphasized that these results are only estimates. Numerous assumptions were made which cannot be verified at this time. Only the general shape and relative proportions, as illustrated by the smoothed curve in the figure, should be considered. Although data were collected to produce a similar histogram for peak (as opposed to average) demand, the figure was not included here because the uncertainties in the peak analysis would be even greater and the results potentially misleading.

4.2.2 Access Needs as a Function of Program Schedule

The level of usage demand on LSS is expected to vary as a function of the licensing process schedule. Demand will be driven primarily by the preparation (i.e., before submittal) and review (i.e., after submittal) of key milestones on that schedule. Estimates of demand variation over time have been made by analyzing this schedule (and correlating staffing level estimates when available). Figure 3 shows the current licensing schedule, reflecting the recent changes resulting from the Nuclear Waste Policy Amendments Act. Note that because of recent changes in the OCRWM program, some of the elements in the figure are in flux, and estimates needed to be used.

Major milestones (general major document submittals) expected to have significant impact on LSS demand have been folded down from these schedules to the time-lines in the middle of Figure 3c. Below these time-lines, a double histogram is presented. Bars above the line represent estimates of the number of users affiliated with NRC, all NRC contractors, States and Indian Tribes and their contractors, intervenors, and legislators. The bars below the line represent estimates of the numbers of users affiliated with DOE and DOE contractors.

No scale has been provided, to avoid over interpretation of the numerical significance of the graph and to uncouple the results presented from the uncertainties associated with estimating the total number of potential users. The graph is intended to portray relative demand over time. The user demand estimates in Section 3.4.12 could be used to scale this graph by taking the combined peak usage estimates to represent the total magnitude between the DOE peak in 1994 (during the preparation of the repository and MRS license applications) and the NRC peak in 1995 (during the review of their review).

Note that LSS usage demand has not been simply correlated to the milestone dates, but accounts for phaseout of detailed technical work in the 3 to 6 months before the delivery date and accounts for initial reading and phase-up during the 3 to 6 months after.

THIS PAGE IS AN OVERSIZED DRAWING OR FIGURE,

THAT CAN BE VIEWED AT THE RECORD TITLED: "TIME-LINE FOR KEY OCRWM ACTIVITIES REPOSITORY SYSTEM", FIGURE 3a

WITHIN THIS PACKAGE

D-01

THIS PAGE IS AN OVERSIZED DRAWING OR FIGURE,

THAT CAN BE VIEWED AT THE RECORD TITLED: "TIME-LINE FOR KEY OCRWM ACTIVITIES TRANSPORTATION SYSTEM", FIGURE 3b

WITHIN THIS PACKAGE

D-02

THIS PAGE IS AN OVERSIZED DRAWING OR FIGURE,

THAT CAN BE VIEWED AT THE RECORD TITLED: "TIME-LINE FOR KEY OCRWM ACTIVITIES MONITORED RETRIEVABLE STORAGE SYSTEM SYSTEM", FIGURE 3c

WITHIN THIS PACKAGE

D-03

5.0 CONCLUSIONS

The background studies described in Section 2 clearly indicate the need for a sophisticated LSS computer system. The size of the system, the variety of the documents that must be included in the data base, requirements for terminals at geographically distributed sites, and simultaneous use of the system, all suggest an advanced computer system that will make use of state-of-the-art technology. The specific and varied demands of users revealed in the interviews reported in Section 3 further indicate the need for a "user-friendly", flexible system that can satisfy the needs of many users with dissimilar needs.

The need for carefully controlled vocabularies for multiple bibliographic and subject data base entry points for structured index searching is a particularly significant finding of the interviews. Many large data bases that have relied on only full-text or full-text in combination with simple headers have proven to produce low recall and precision. The experience of the potential users interviewed for this report underscore the findings of the literature. Their experience and needs (the best proxy available for LSS usage prediction) indicates the necessity for comprehensive, in-depth cataloging of LSS documents.

This needs analysis has resulted in the identification of numerous requirements which must be met by the LSS, including those currently identified by NRC as needed to meet the 3-year repository licensing processes. In some cases the evidence is firm. In other cases, the evidence is preliminary and indicates that further analysis and cost/benefit studies must be conducted to determine if the apparent need is sufficient to support a clear system requirements. Those requirements which clearly fall into the firm category are:

General Functions

- The LSS should include capabilities for managing various types of data bases (document, tracking data, etc).
- Information should be stored in the form of headers for all records and in full-text for many, if not all documents.
- The system should include the capability for efficient and accurate data retrieval using a variety of methods.
- Data should be capable of being distributed in hardcopy form.
- The system should be capable of generating various types of reports.
- The system should be easy to use with minimum training necessary, containing built-in help functions, and providing assistance when needed, either through an expert system or on-call assistance.
- There should be a procedure to identify and minimize or avoid duplicate records.
- The records should be maintained in a secure environment.
- Electronic mail capability should be provided.

- The data should be entered and maintained under an independently verified quality assurance program.

Data Retrieval Methods

- Structured index searching *via* detailed and extensive headers should be available, involving subject terms and keywords assigned with the aid of a controlled vocabulary.
- Full-text search capability on both document text and headers should be available.
- Search aids including thesaurus, boolean logic, and proximity searches should be available to meet the performance requirements identified.

Data Bases

- The system should include data bases for documents, regulations, tracking of issues and commitments, and indexes for non-documents such as physical sample inventories.

User Community

- The number of users needing the system will probably exceed 350 at peak demand.
- Major geographic centers of users will be Washington D.C., Las Vegas, Nevada and San Antonio, Texas. Other locations, comprising about 10% of the users, are expected.
- Discernible usage patterns will exist in terms of user characteristics, time-of-day use, and use in relation to the program schedule.

Performance

- The system should be designed to maximize recall and precision with minimum performance targets on the order of 80 percent.
- Response to large queries should be available in approximately 15 to 30 minutes.
- Documents should be available for viewing on a screen in less than 30 seconds after they are identified.
- Routinely, hardcopy of documents should be available overnight.

Schedule

- The system should be available for use by August of 1990, which would provide over 4 years of use prior to license submittal.

Certain additional needs were identified in the course of this study, however additional study will be required to determine if they will become firm requirements. These include:

General Functions

- Electronic (bit-map) images may be required for some records, particularly graphs, diagrams and maps.

- The capability for downloading data may be desirable.
- Some files (such as privileged information or annotated records) may require limited access or privacy.

Data Retrieval Methods

- The utility of abstracts, particularly if they must be created for LSS cataloging, is not clear, especially for documents which can be searched in full text.
- The utility of misspelling tolerance is also unclear, but should be selectable if included.

User Community

- Some users have expressed interest in a priority access system.

The conclusions and requirements derived from this report represent a set of findings to be revised as the requirements definition progresses. These refinements are not expected to significantly alter the conclusions, but rather to allow them to reflect in detail any statutory changes and to refine the basis for quantitative system specifications. The activities which are expected to contribute to this process over the next 6 months include:

- The Preliminary Data Scope Analysis (the second report of this series) will be based in part on interviews to be conducted with subject experts to help define the information to be included in the LSS data base. Interviews of this nature will not be limited to data scope but must overlap into data retrieval and functional definition of the LSS. For example, the potential requirement for electronic storage and retrieval of graphic information, diagrams, and maps will be better defined when the quantity, characteristics, and use of such information is identified in the data scope analysis.

- The Negotiated Rulemaking Advisory Committee will continue deliberations on the rulemaking process which will identify, among other things, their judgment on the information to be included within LSS, the types of information to be in full-text, the need for privacy of files, input to the data entry process, and requirements for access to LSS. As the conclusions of this committee become firm, they will be incorporated into the refinements of these requirements.

- The resultant Conceptual Design Analysis (the third report in this section) will be a specific concept which can be used as a reference for detailed review. Some of the questions used in the interviews thus far have, of necessity, been general and without benefit of such a frame of reference. The conceptual design analysis will support a detailed scrutiny of requirements, rather than continuing to investigate general concepts.

- An important aspect of the system definition and design is the development of an LSS prototype which will provide the technical data required to refine and complete system specifications. The prototype will provide significant insight into data entry and retrieval techniques. The use of a prototype system, with a representative data base and truly prototypical software, will be tested by potential users. These tests will provide firm technical information to support refinements in retrieval methods and header fields.

As shown by the series of documents which have preceded this report, LSS requirements have been extensively reviewed, considered and discussed over a period of years. Through the process outlined in this report, the subject is now receiving the concentrated study and analysis necessary to create specifications from which a system can be designed. This could not have been achieved without the assistance of all involved parties and we look forward to continued cooperation.

REFERENCES

Aerospace, 1986; Requirements Definition for Licensing Information Management System for Nuclear Waste, Draft Revision 2, the Aerospace Corporation, prepared for NRC, March 1986.

ANSI/ASME NQA-1, 1983; Quality Assurance Program Requirements for Nuclear Facilities, American National Standards Institute and American Society of Mechanical Engineers, 1 July 1983.

Browning, R.E., 1985; Shortening NRC Review Time for the Repository Authorization Application, Nuclear Regulatory Commission, 9 May 1985.

DOE, 1986; Quality Assurance Plan for High-Level Radioactive Waste Repositories, DOE OCRWM Office of Geologic Repositories, August 1986.

DOE, 1987; Request for Proposal: Design and Implementation of a Licensing Support System, DOE Office of Civilian Radioactive Waste Management, 11 February 1987.

Federal Register, 1986, 1987; 18 December 1986 (51 FR 45338) and 5 August 1987 (52 FR 29024).

Jordan, 1986; Discovery and rulemaking perspectives on the use of an information storage and retrieval system in the licensing proceedings for the high-level waste repository, John S. Jordan & Associates, prepared for NRC, 14 February 1986.

NARA, 1987; Electronic Recordkeeping, National Archives and Records Administration Bulletin 87-5, 11 February 1987.

NRC/DOE, 1986; Agreement in Principle, Davis/Rusche, February 1986.

Olmstead, W.J., 1986; Regulatory Issues Associated with Licensing A High-Level Waste Repository, proceedings of the Symposium on Waste Management, March 1986.

SAIC, 1986; NNWSI Project Information Management System Concepts Evaluation Report, Science Application International Corp., prepared for OCRWM NNWSI, August 1986.

SAIC, 1988; Information Management System Bridge Program Preliminary Test Results, Science Applications International Corp., prepared for OCRWM NNWSI, January 1988.

Weston, 1985; Licensing Information System Requirements Study: Draft Report of Findings, Roy F. Weston Inc., prepared for OCRWM Office of Geologic Repositories, 2 October 1985.

Young, 1987; LSS Functional Requirements and Design Concept Report, Arthur Young International, prepared for OCRWM, 31 March 1987.

APPENDIX A
List of People Interviewed

Charlotte Abrams
Division of High Level Waste Management
U.S. Nuclear Regulatory Commission

Avi Bender
Administration and Resources Management Office
U.S. Nuclear Regulatory Commission

Felton Bingham
Environmental Assessment Division
Sandia National Laboratories

Charles (Ched) Bradley
Office of NEPA Project Assistance
U.S. Department of Energy

Francis (Chip) Cameron
Office of the General Counsel
U.S. Nuclear Regulatory Commission

Richard Codell
Division of High-Level Waste Management
U.S. Nuclear Regulatory Commission

Robert Cranwell
Manager, Waste Management Systems Division
Sandia National Laboratories

Helen Cummings
Energy Library (Germantown)
U.S. Department of Energy

Mark Delligatti
Division of High-Level Waste Management
U.S. Nuclear Regulatory Commission

Norm Eisenberg
OCRWM Office of Geologic Repositories
U.S. Department of Energy

Stan Echols
Office of the General Counsel
U.S. Department of Energy

Dan Fehringer
Division of High-Level Waste Management
U.S. Nuclear Regulatory Commission

APPENDIX A
List of People Interviewed
(Continued)

Steve Frank
Office of NEPA Project Assistance
U.S. Department of Energy

Mark Frei
OCRWM Office of Geologic Repositories
U.S. Department of Energy

Mike Glora
Project Regulatory Compliance Department
Science Applications International Corporation

Donna Hickling
Manager, Marketing Resource Center
Booz, Allen & Hamilton

Ginger King
OCRWM Office of Policy and Outreach
U.S. Department of Energy

Hannah King
Energy Library (Germantown)
U.S. Department of Energy

Melinda Kassen
Attorney
Environmental Defense Fund

John Linehan
Division of High-Level Waste Management
U.S. Nuclear Regulatory Commission

Mike Martin
OCRWM Records Management Facility
KOH, Inc.

Peter McGrath
Group Manager, Energy Systems Group
Science Applications International Corporation

Abigail Muller
Central Research Library
Oak Ridge National Laboratories

Mal Murphy
Special Deputy Attorney General
State of Nevada

APPENDIX A
List of People Interviewed
(Continued)

Gerald Parker
OCRWM Office of Geologic Repositories
U.S. Department of Energy

Carol Peabody
OCRWM Office of Geologic Repositories
U.S. Department of Energy

Jerome Pearring
Division of High-Level Waste Repositories
U.S. Nuclear Regulatory Commission

Jeanne Perrone
Energy Library (Germantown)
U.S. Department of Energy

John Randall
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission

Ed Regnier
OCRWM Office of Geologic Repositories
U.S. Department of Energy

Kathleen Ruhlman
Public Document Room
U.S. Nuclear Regulatory Commission

Len Scoplar
Manager of Regulatory, Safety & QA Department
Roy F. Weston, Inc.

Trudy Scott
Energy Library (Washington, D.C.)
U.S. Department of Energy

Jaynie Shaheen
OCRWM Office of Policy and Outreach
U.S. Department of Energy

Betsy Shelburne
Public Document Room
U.S. Nuclear Regulatory Commission

Barry Smith
Attorney
Tourtellotte, Ross and Gray

APPENDIX A
List of People Interviewed
(Continued)

Jan Statler
Manager, OCRWM Records Management Facility
KOH, Inc.

Ralph Stein
Director, OCRWM Office of Engineering & Geotechnology
U.S. Department of Energy

Eileen Tana
Division of High-Level Waste Management
U.S. Nuclear Regulatory Commission

Michael Teubner
Lockheed Engineering and Management Service Company

Dean Tousley
Attorney
Harmon and Weiss

John Trapp
Division of High-Level Waste Management
U.S. Nuclear Regulatory Commission

Mary Louise Wagner
Staff, Energy and Natural Resources Committee
U.S. Senate

Jeff Williams
Office of NEPA Project Assistance
U.S. Department of Energy

APPENDIX B
Abbreviations Used

ANSI	American National Standards Institute
ARS	Automated Records System
ASCII	American Standard Code for Information Interchange
ASLB	NRC Atomic Safety and Licensing Board
ASME	American Society of Mechanical Engineers
BASIS	A text-oriented data management system developed by Battelle
CDC	Control Data Corporation
CFR	Code of Federal Regulations
DEIS	Draft Environmental Impact Statement
DM	A relational data base system developed by Battelle
DOE	Department of Energy
EIS	Environmental Impact Statement
FEIS	Final Environmental Impact Statement
FFRDC	Federally Funded Research and Development Center
FR	Federal Register
GSA	General Services Administration
IMS	Information Management System
IMSBP	Information Management System Bridge Program
IS&R	Information Storage and Retrieval
JURIS	US Department of Justice Legal Information Retrieval System
LA	License Application
LEXIS	Mead Data Central Legal Data Base
LIMS	Licensing Information Management System
LIS	Licensing Information System

APPENDIX B
Abbreviations Used
(Continued)

LSS	Licensing Support System
M&O	Management and Operations contractor
MRS	Monitored Retrievable Storage
MT	Metric Tons
NARA	National Archives Records Administration
NBS	National Bureau of Standards
NFPA	National Fire Protection Association
NNWSI	Nevada Nuclear Waste Storage Investigations
NRAC	Negotiated Rulemaking Advisory Committee, officially known as the HLW Licensing Support System Advisory Committee
NRC	Nuclear Regulatory Commission
NTS	Nevada Test Site
NWPA	Nuclear Waste Policy Act of 1982
OCR	Optical Character Recognition
OCRWM	DOE Office of Civilian Radioactive Waste Management
ODS	Optical Disk Storage
OGR	OCRWM Office of Geologic Repositories
OMB	Office of Management and Budget
QA	Quality Assurance
RFP	Request For Proposal
RIS	OCRWM Salt Repository Project Records Information System
SAIC	Science Applications International Corporation
SARP	Safety Analysis Report Package
TFS	Technical Field Services
WBS	Work Breakdown Structure
WP	Waste Package

LICENSING SUPPORT SYSTEM
PRELIMINARY DATA SCOPE ANALYSIS

SCIENCE APPLICATIONS INTERNATIONAL CORPORATION

TABLE OF CONTENTS
(continued)

TABLE OF CONTENTS

PREFACE

This is the second in a series of four reports on the Licensing Support System (LSS) prepared by the DOE Office of Civilian Radioactive Waste Management (OCRWM). The LSS is an information management system intended to support the needs of all the parties involved in repository licensing, including the Department of Energy (DOE) and the Nuclear Regulatory Commission (NRC). The reports in this series are:

- Preliminary Needs Analysis
- Preliminary Data Scope Analysis
- Conceptual Design Analysis
- Benefit-Cost Analysis.

The Preliminary Data Scope Analysis, presented in this report, and the Preliminary Needs Analysis, issued in February 1988, constitute the system requirements basis for developing a conceptual LSS design, which will be presented in the third report. The Benefit-Cost Analysis will present an evaluation of alternatives within this conceptual design. These four reports, and subsequent refinements, are intended to provide the basis for determining the LSS design specifications.

1.0 INTRODUCTION

1.1 Purpose and Scope

The purpose of this analysis is to determine the content and scope of the Licensing Support System (LSS) data base. Both user needs and currently available data bases that, at least in part, address those needs have been analyzed. This analysis, together with the Preliminary Needs Analysis (DOE, 1988d) is a first effort under the LSS Design and Implementation Contract toward developing a sound requirements foundation for subsequent design work. These reports are preliminary. Further refinements must be made before requirements can be specified in sufficient detail to provide a basis for suitably specific system specifications.

There have been a number of previous examinations of the LSS data scope requirements. These are summarized in Section 2. In Section 3.1 a description and schedule of the DOE Office of Civilian Radioactive Waste Management (OCRWM) high level waste repository program is given with an emphasis on the impact of information requirements on LSS. A summary of forthcoming site characterization and licensing documents is also provided. This is followed, in Section 3.2, by profiles of current data bases containing information relevant to LSS, including OCRWM and NRC collections of documents. The required contents of the LSS is then addressed in Section 3.3 from the perspective of potential user needs. This section uses the findings of the Preliminary Needs Analysis as a basis for translating the needs into specific information requirements. The information to be included in the LSS Regulations Access Subsystem is then examined in detail. An estimation of the size of the data base required by LSS users in August 1990 (when LSS is partially loaded and available) is presented in Section 4. Projections from that date are made for the next twenty years. The report concludes with a brief summary of major findings.

This document provides a baseline for what is known at this time. Additional analyses, currently being conducted, will provide more precise information on the content and scope of the LSS data base.

1.2 Background

The evolution of requirements for LSS is discussed in detail in the Preliminary Needs Analysis (DOE, 1988d). That analysis concluded that LSS must:

(1) Serve as the sole basis for expedited document discovery

(2) Provide access to licensing information so that all parties' legal counsel and their experts can address the grounds for repository licensing decisions and determine the soundness of technical work

(3) Provide an automated library of reports and other bibliographic

materials of use to OCRWM and NRC technical staff in conducting their work on licensing document development and review

(4) Serve as a mechanism for tracking OCRWM compliance with repository licensing regulations.

An overview of the structural components of LSS is given in Table 1. As conceived in the RFP (DOE,1987a), the LSS computer system would include four application software subsystems: a Regulations Access subsystem, a Records Access Subsystem, an Issue Tracking Subsystem and a Commitment Tracking Subsystem. Table 1 provides a brief description of the subsystems and indicates the general size of each of the subsystem data bases.

The analysis of LSS requirements presented in the Preliminary Needs Analysis does not explicitly focus on the scope of information to be included in the LSS data base, nor on the amount of information that data base represents. The following section has been included to summarize previous treatment of these data scope issues.

2.0 PREVIOUS EXAMINATIONS OF THE INFORMATION TO BE INCLUDED IN A LICENSING SUPPORT SYSTEM

This section summarizes information about the content and size of a licensing support system as discussed in previous analyses for the DOE, the NRC, and in other documents relevant to the high-level radioactive waste program. These documents are valuable in that they present information on subject content, types of documents or records, page counts, numbers of records, or similar data that can be used in determining the data requirements for the LSS.

2.1 Nuclear Waste Policy Act (NWPA) of 1982

The NWPA does not specifically mention an information system to support the licensing process, and therefore it does not specify the size or content of such a system. However, it is clear that all documents explicitly required by the Act should be included in such a system, and will be needed by LSS users. These include specific documents (such as the Environmental Assessment, the Site Characterization Plan, and the Environmental Impact Statement) and general documents (such as reports, petitions, hearings, transcriptions, correspondence, and studies from DOE, NRC, the states involved, the President, Congress, and other peripheral agencies). A precise list of the documents required by the Act cannot be made, since some documents are produced annually, some reference supporting documents are not explicitly mentioned in the Act, and yet other documents require the reporting of data collection or testing or interim reporting before a final document is produced. The Act is, however, a good source for identifying the major documents and categorizing document types that will be generated.

Table 1. SALIENT CHARACTERISTICS OF THE LICENSING SUPPORT SYSTEM

SUBSYSTEM	DESCRIPTION	SIZE OF DATABASE/USERS
Regulations Access Subsystem	A text storage subsystem, on-line containing the full text of documents that impose legal requirements on the construction or operation of the geologic repository.	300 documents, 100,000 pages. Used by design and licensing engineers.
Records Access Subsystem	Full text on-line with index to archive files. Contains text of documents subject to discovery during licensing hearings and appeals.	Millions of documents, tens of millions of pages. Used by attorneys and licensing staff.
Issue Tracking and Commitments Tracking Subsystems	Issues are efforts that clarify what a part of a regulation requires or how compliance will be achieved. Commitments are agreements between organizations to perform an activity, adhere to a standard, etc. Details re source, organizations, regulations, milestones with linkages to regulations, documents, etc.	Up to 4,000 issues and 4,000 commitments. Used by DOE, NRC technical staff.

2.2 Mission Plan for the Civilian Radioactive Waste Management Program

Part II of the Mission Plan presents the information needs of the Program as required by the NWPA. It begins with a discussion of the four major issues, the Plan and then details the information necessary to resolve not only the key issues, but also the subordinate issues that make up the issues hierarchy. The issues hierarchy consists of three levels of detail: key issues, which are broad questions derived from the siting guidelines and have as their concern the suitability of the site in terms of protecting the public and the environment; issues, which are questions whose answers will provide the information to resolve the key issue; and information needs, which are specific information needed to answer the questions posed in the issues. Some of these information needs can be provided by existing documents; other information will be obtained through data collection and analysis. As an example, Key Issue 3 concerns protecting the quality of the environment throughout the entire disposal process (from repository siting through decommissioning and transportation operations), without causing unacceptable risks to public health or safety. An attendant issue asks if the site can be located to protect the environment and if significant adverse environmental impacts can be mitigated by reasonable measures. Information needs to determine the environmental conditions at the site include such concerns as existing air-quality levels and trends, existing surface- and ground-water trends, soil characteristics, land use patterns and trends, and noise levels.

In addition, potential financial, political, legal, and institutional problems are recognized and plans for resolving these problems are discussed. An example of such a problem is ligitation by states, tribes, or other parties that would prevent the DOE from meeting the NWPA schedule. In addition to consultation and cooperation agreements with these parties, the DOE plans to provide them with access to program information and also plans to document all key decisions and program actions to the level of detail needed to survive a legal challenge.

The resolution of these issues and problems will produce documents and non-document items, expanding the document base established by the NWPA. The Mission Plan also provides a rich source of information on the subject material needed in the preparation of repository licensing documents, during licensing and in the review of documents supporting the repository license application, and therefore on the subject material to be included in the LSS.

2.3 Licensing Information System Requirements Study

In 1985, Roy F. Weston, Inc. produced a report (Weston, 1985) on the requirements of a system to meet the information needs of those parties involved in the repository program. This report examines not only the question of what information is to be included in such an information system, but also what information is to be excluded. The information system requirements include all OCRWM-related regulations and regulatory guidance; the licensing schedule network; all issue-related plans, activities, and resolutions; all commitments-related activities and results; key document preparations and modifications; and all OCRWM-produced or received information pertaining to regulatory compliance. Therefore, the study indicates, the system must contain not only information on completed activities, but also planned and continuing activities. The type of documents to be excluded are personnel records, drafts documents, duplicate copies, etc. Key documents for inclusion are defined as "any program document specifically required by NWPA, Federal, State, or local regulations that is either prepared by the DOE or by a regulatory agency". Information in the system would be maintained through the decommissioning of the repository.

2.4 Discovery and Rulemaking Perspective on the Use of an Information Storage and Retrieval System in the Licensing Proceedings for the High-Level Waste Repository

The relevance of the information entered into a licensing support data base and the completeness or reliability of the entire data base, i.e. is all the relevant material included are discussed in a report for NRC by John Jordan & Associates (Jordan, 1986). The criteria for relevance are well defined in law and for LSS purposes include: (1) pertinence to the contentions defined by the licensing board and (2) pertinence of the subject matter of those contentions. Further constraints are placed on information to be included by privilege, confidentiality and Freedom of Information Act (FOIA) exemptions. Other considerations that might further define the content of the data base will come from the NRC Negotiated Rulemaking Advisory Committee and from identification by DOE of issues or problems that could arise during the repository licensing process.

2.5 Requirements Definition for a Licensing Information Management System for Nuclear Waste

A report for the NRC by The Aerospace Corporation (Aerospace, 1986) also discusses the content of a licensing information system. The proposed criterion for inclusion of information in the system is "any record likely to be requested that pertains to high-level waste in compliance with 10CFR2 (Rules of Practice for Domestic Licensing Proceedings)." This report includes estimates of the amount of information that would be contained in such a data base. A volume of 300,000 documents is assumed to be in the system by 1999. According to this report, documents not subject to discovery are to be excluded, e.g., handwritten drafts and drafts of published final material.

However, the report notes, materials other than legally-required documents are not necessarily excluded from the data base. Indeed, it is recognized that the system should include other material, but decisions need to be made as to what this material would be. If other materials are included, however, material that meets legal requirements should be distinguished in some manner from that which does not.

Other recommendations from this report that would influence the size of the data base include the storage of an embedded thesaurus, thorough indexing of each document, references to preceding and subsequent documents, document reviews, and FOIA requests for documents.

2.6 Quality Assurance Plan for High-Level Radioactive Waste Repositories

According to this document (DOE, 1986d) the LSS is to be used "as the repository for and custodian of all OCRWM Office of Geologic Repositories (OGR) Quality records". The OGR will not establish a separate quality records system but will rely on the LSS. Therefore, the LSS must meet the quality assurance requirements established by information from the ANSI/ASME NQA-1. This document defines a QA record as "a completed document that furnishes evidence of the quality of items and/or activities affecting quality". To meet these and other similar requirements, OGR has set the following criteria for records that are to be retained in the LSS data base for the life of the repository:

- records which "may be used in repository licensing",
- records which demonstrate the "capability for safe operation",
- records which would be of value in "maintaining, reworking, repairing replacing or modifying an item",
- records which would have value in "determining the cause of an accident or malfunction of an item",
- records which provide "required baseline data for in-service inspections".

To reduce the volume of retained records, records required to show compliance with a requirement that does not meet one or more of the above criteria will be considered as nonpermanent records and may be removed after the retention period set for that record has elapsed. Prior to removing the record from the system a review would be conducted to ensure that the record is no longer required. An index to all records is to be maintained, the plan indicates, and the index will track revisions of documents in the system. Records are either documents or items, e.g., physical samples, magnetic material. Revisions will be filed as separate documents.

2.7 NNWSI Project Information Management System Concepts Evaluation Report

The content and size of a licensing support data base is also discussed in this report (SAIC, 1986a). Regulatory documents considered to be necessary for inclusion are NRC requirements, DOE siting guidelines and Environmental Protection Agency (EPA) guidelines. The report notes that in order to show compliance with these requirements, compliance issues need to be identified, analyzed, tracked, and closed. Each step from identification to closure of an issue should be included in the data base. Given that issues are not all initially defined and that documents already exist that may support future issues, the approach taken is to consider all relevant project material potentially discoverable. This information is generated by three groups: (1) those originating project-related documentation within the organization of the project participant, (2) those within a non-participant organization who are transmitting project-related information to a project participant, and (3) those within a non-project group originating or receiving project-related documentation within the organization of the project participant. The latter group of documents are the most difficult to identify and obtain since the involvement of these people in the project is peripheral. This includes documents concerning corporate quality assurance, purchasing, etc.

The report concludes that over 6 million documents are expected to meet the above requirement by the year 2000. Based upon an assumption of four pages per document up to 1991 and four pages per document for off-site contractors and two pages per document for on-site contractors from 1991-2000, the total pages required to be contained in the LSS by the year 2000 would be approximately 18 million according to this evaluation.

2.8 Supporting Data and Calculations for the NNWSI Project Information System Concepts Evaluation Report

This report (SAIC, 1986b) provides supporting data for the document discussed in Section 2.7 above. It, like the document it supports, is concerned only with the Nevada Nuclear Waste Storage Investigations (NNWSI) Project. The document and pages estimates it provides are based on a survey that asked project participants and NNWSI contractors for estimates of the amount of Project information to be generated in the 10-year period from 1980 to 1990. Responses from each participant or contractor show document estimates by project activity, document type, document count, pages per document, and total pages.

Representative disciplines and activities covered by this report include geology, hydrology, waste package environment, testing, drilling, project control, and quality assurance. Representative document types include field notebooks, geological maps, aerial photos, core logs, technical reports, computer models, personnel qualifications, drawings, specifications, budgets, and schedules. Document and page count projections for 1990 for all types of documents originated by the nine project participants at this site were 1,039,236 and 3,944,157, respectively.

2.9 LSS Functional Requirements and Design Concepts Report

This study by Arthur Young International (Young, 1987) not only details material for inclusion and exclusion, but also discusses further constraints on submitted material. Their approach is that the primary purpose of the LSS is to support information needs during the licensing of the high-level radioactive waste repository. According to this report, support of other phases of the repository program would initially not be included in this system, although the system could and should be expanded to include these at a later date. Records to be included are those received by DOE or its supporting organizations from an outside group, agency, or individual. A record originating within DOE or its contractors is to be included at the stage when it is ready for formal distribution either as a finished product or as a draft for review, it is cited in support of another submitted record. Materials not to be included in the system, the report notes, are non-record material as specified in DOE Order 1324.2; financial records unrelated to the health, safety or environmental impacts of the repository; personnel records not required by a QA program; attorney work packages; attorney-client privileged records; unrelated internal organizational memoranda; items such as electronic mail, telephone messages, etc. (if these items contain information of value, they are to be reformatted as memoranda and submitted as such); and drafts of internal correspondence or drafts and marked-up copies of documents not distributed outside the originating office. Types of material include all revisions of related documents, correspondence, inquiries from interested parties, related computer software, photographs, maps, drawings, QA records, regulations, standards, procurement documents, schedule and budget reports. The LSS must also record and track issues and commitments.

In addition to the type of material to be included or excluded, other constraints are placed on information in the system according to this report. These constraints include verification that the information has been authorized and certified for entry, assurance that the records are accurately captured and displayed by the system, verification that the material is not a duplicate of an existing record, and assurance that no changes are made to the record per se after it is stored in the data base. Changes or modifications may be made to identifying elements attached to each record. These elements include keywords, header information, abstracts, etc. Such elements are an additional source of material to be added to the system.

2.10 Basic Ideas Underlying Design and Development of an Information System to Support Proceedings before the NRC on a High-Level Waste Repository

This document by John S. Jordan & Associates (Jordan, 1988) distinguishes between data entered into the existing DOE Automated Records System (ARS) and the future LSS. Distinctions between the two systems include the following:

ARS	LSS
Documents originated or received by OCRWM and its contractors	Documents originated or received by by any party
Any subject	All materials related to repository licensing
Transmittal letters, acknowledgments	Only documents relevant to licensing
No handwritten notes, preliminary drafts, buck sheets, etc.	Any type of document that is relevant and goes through document control
Support repository until final closing	Support only the licensing process

ARS obviously has much candidate information for the LSS, but each document must satisfy LSS requirements before it can be included in the LSS.

Issues to be considered that will affect the content of the LSS include defining "relevant" documents; identifying sources of such documents; how to treat privileged documents; and how to identify, acquire, and store the documents (estimated as being a minimum of 3 million after elimination of the Texas and Washington sites).

3.0 CHARACTERISTICS OF THE LSS DATA BASE

This section examines a number of key elements that will determine the content of LSS. Most important of these is the OCRWM program and milestone schedule. Documents generated by and associated with the program are an obvious must for LSS. The OCRWM and NRC data bases now in existence are a reflection of current information needs and, thus, must be considered as a major source of LSS materials. This section also describes in detail the regulations data base seen as the core of the LSS Regulations Access Subsystem.

3.1 The OCRWM Program

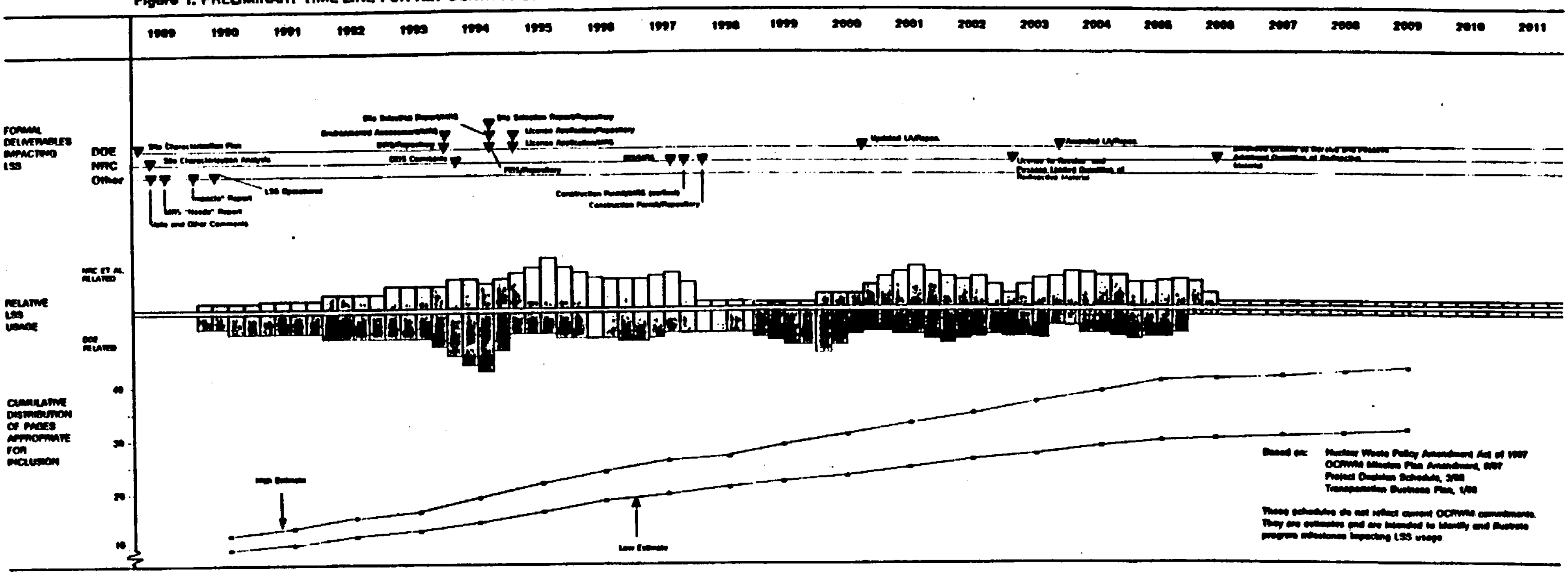
The Preliminary Needs Analysis shows that the vast majority of requirements on the LSS arise from or are related to the major milestone processes. This section summarizes the major program elements of the OCRWM program, including the issue resolution process, and identifies and characterizes the key program documents that will be generated over time. It should be recognized that both DOE and NRC are still in the preliminary stages of defining the flow of documents and their official classifications as well as program milestones. The documents discussed here are based on preliminary planning by DOE and NRC. While the basic document flow is not likely to change appreciably, the official classifications of documents may change.

3.1.1 Overall Program Schedule

Figure 1, adapted from the Preliminary Needs Analysis, shows a preliminary timeline depicting the major program activities, as defined at this time, and estimated schedules for the OCRWM program that directly affect the LSS. The timeline encompasses the repository program and the transportation and monitored retrievable storage programs. However, most of the discussion will center on the repository program because the bulk of documents to be stored in the LSS will result from repository activities. It is important to note that the schedule is tentative and that the program is still evolving.

Using this timeline, which also identifies the major licensing-related documents, the Preliminary Needs Analysis has estimated the relative usage of the LSS over time. The peaks correspond primarily to the stages when a license application, or amendment to the license application, is being developed by DOE and reviewed in a formal licensing proceeding by NRC, the potential host State and other parties. These peaks occur from 1994 to 1997 and again from 1999 to 2004.

Figure 1. PRELIMINARY TIME-LINE FOR KEY OCRWM ACTIVITIES



3.1.2 Overview of the Major Program Phases Impacting LSS

The activities depicted on the timeline can be grouped into four major phases from now until the early 2000s. They are: (1) site characterization, (2) licensing, (3) construction, and (4) operation. This report will focus primarily on the first two phases, site characterization and licensing, since most of the licensing-related documents will be generated and accessed during these two phases.

3.1.2.1 The Site Characterization Phase

The first major phase, site characterization, will continue until 1994. Site characterization is a program of studies directed at collecting the information necessary to: (1) demonstrate the suitability of a site for development as a repository, (2) design the repository and waste package, and (3) demonstrate compliance with all regulatory requirements pertaining to public health and safety. The scope and content of the site characterization program must produce the data and information necessary to satisfy the requirements under 10 CFR 960 (General Guidelines for the Recommendation of Sites for Nuclear Waste Repositories; Final Siting Guidelines). This regulation and its references to other regulations, contain all the technical criteria for which compliance must be demonstrated. (DOE, 1988a) Based on these requirements, a common set of issues has been developed and are contained in the DOE "Issues Hierarchy for a Mined Geologic Disposal System" (DOE, 1987b). To satisfy the regulatory requirements, DOE must resolve all issues defined in the hierarchy, which will be accomplished primarily during the site characterization phase. Figure 2 contains DOE's 12-step issue resolution strategy (DOE, 1987b).

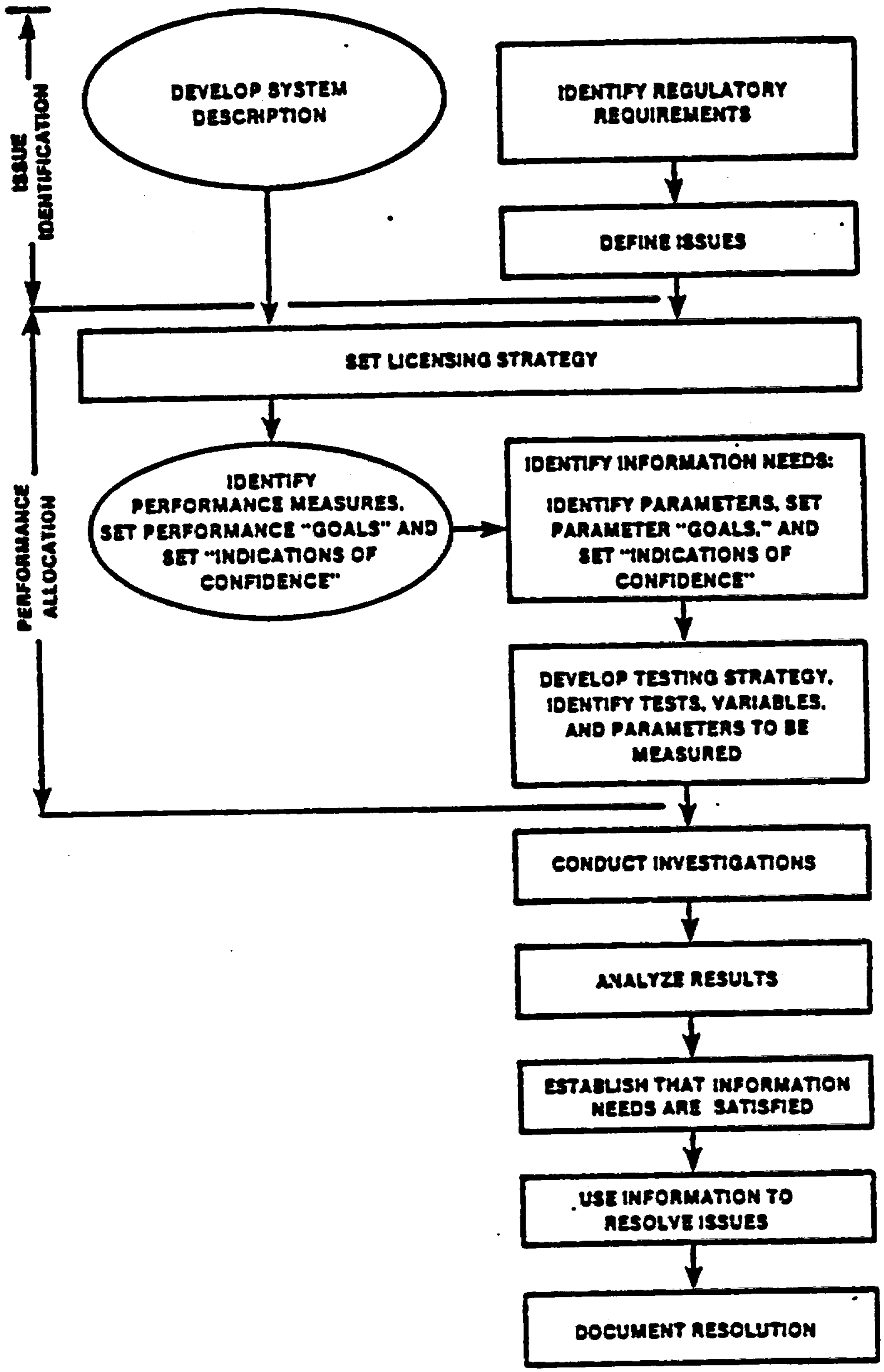
The issues hierarchy consists of three levels of detail: key issues, issues, and information needs. Key issues are broad-level requirements of overall site suitability which relate to general regulatory objectives. Issues are narrower questions, subordinate to a key issue, that address more specific regulatory requirements pertaining to the site, features of the design, and performance of the system. (Resolution of all issues that support a key issue results in resolution of that key issue.) Information needs represent the information required to resolve the issues and thus provide the basis for the studies, tests and analyses that will take place during the site characterization phase. (DOE, 1988a)

Issuance by DOE of the Site Characterization Plan (SCP) is the first major milestone of the site characterization phase. The basic purpose of the SCP is threefold:

(1) To describe the site, the preliminary designs of the repository and waste package, and the waste-emplacement environment in sufficient detail so that the basis for the site characterization program can be understood;

(2) To identify the issues to be resolved during the site characterization, to identify the information needed to resolve the issues, and to present the strategy for resolving the issues; and

Figure 2.
THE DOE RESOLUTION STRATEGY



(3) To describe general plans for the work needed to resolve outstanding issues. (DOE, 1988c)

A consultative draft SCP (CDSCP) was issued to the NRC and State of Nevada for review and comment in January 1988. Following review of NRC and State comments, DOE expects to issue the final SCP in late 1988, which will be followed by more detailed Study Plans for NRC comment. NRC will issue its Site Characterization Analysis in mid-1989. In 1989, DOE expects to start exploratory shaft construction.

For the next six or seven years, DOE will conduct an extensive program of site characterization, which will consist of surface-based field studies and other tests. During this time period, DOE will be collecting the data necessary to satisfy the information needs for each issue identified in the issues hierarchy and in the SCP and will generate a multitude of reports to document the data obtained and to demonstrate how the data supports the various issues. In addition, a series of workshops, meetings, data reviews and other interactions will take place between DOE, NRC and the State in an attempt to reach early resolution on as many licensing issues as possible before the actual licensing hearing. The reports produced during this phase will also contribute to the development of the repository and waste package designs, the environmental impact statement, the recommendation report to the President, and the license application.

3.1.2.2 The Licensing Phase

In accordance with the Nuclear Waste Policy Act, as amended, NRC must complete its licensing review and hearings within three years of receipt and acceptance of DOE's license application. (A one-year extension is possible with Congressional approval.) DOE expects to submit the license application containing its application for a construction authorization to NRC in early 1995. NRC has no precedent for completion of a major licensing proceeding in three years (Jordan, 1986). Experience with highly contested reactor proceedings indicate that five to seven years is not an uncommon expectation (Olmstead, 1987). For reactor cases, the license application typically comprises 10 or more large volumes of material consisting of both safety and environmental factors (NRC, 1987). The repository application will likely be much greater in length than that required in a commercial nuclear reactor case.

The license application for the repository will consist primarily of the Preliminary Safety Analysis Report and the Final Environmental Impact Statement, both of which will contain the repository and waste package designs, and other related information. These major documents will contain DOE's data, analyses and proposed demonstrations of regulatory compliance based on information derived from the site characterization activities.

After formal docketing of the license application with the NRC, an Atomic Safety Licensing Board (ASLB) will convene and conduct a prehearing conference with the involved parties, including any interveners. The issues and contentions to be pursued in the licensing proceedings are defined during this prehearing conference by the ASLB. After this point, several activities affecting data scope and content will proceed in parallel.

First, the parties to the licensing proceeding begin a formal discovery process (see Section 3.3.1.6). This involves fact finding to obtain information from DOE, NRC, and others in order to develop challenges to the adequacy of the technical data and analyses presented. While informal discovery (e.g., consultative exchanges of information) will have taken place during the several years prior to filing of the license application, formal discovery is generally directed to the contentions accepted by the ASLB. For a proceeding of this nature (first of its kind and multiple well-funded parties) the traditional discovery process would involve hundreds to thousands of requests for information, the handling of thousands of documents, and the filing of multiple interrogatories, depositions, affidavits and testimony. Significant time and resources would be spent in requesting, searching, retrieving, developing, copying and mailing these massive quantities of documents. NRC staff recently estimated that,under existing rules, a document production request in a large case can require 12 to 18 months of manual effort. Large file rooms have to be established by each party and time is needed to manually sort and select records at the site of production. With multiple well-funded parties this means extensive travel, scheduling, review and motion practice (Olmstead, 1987). For reactor cases, the formal discovery phase has typically taken 12 to 24 months (Jordan, 1986). However, as characterized by the Chairman of NRC's ASLB, the repository case may well be the largest administrative proceeding ever conducted. He estimates that the number of fully-funded parties to the licensing hearing will be 10 to 30 times greater than that in reactor licensing cases and that the number of documents subject to the discovery phase will be about 30 to 40 times larger (Cotter, 1986).

With only three years under law to complete the proceeding, it is obvious that steps must be taken to reduce the time required for formal discovery. To cope with this difficult administrative problem and thus avoid substantial delays in the proceeding, NRC proposed the development of a data base management system (the LSS) that would be on-line and available to all parties to the proceeding for discovery. Document discovery, both informal and formal, would take place within the system, which could save significant time and professional resources. If the data base were available well in advance of the start of the proceeding, the formal document discovery procedures could be reduced to a few months (Olmstead, 1987). An NRC negotiated rulemaking, involving all potential parties to the repository proceeding is underway and will set the structural and operational parameters of the LSS.

Second, the NRC staff will conduct its own safety and environmental reviews of the license application. Prior to this time and throughout the proceeding, the NRC and DOE staffs, as well as interested parties, will continually interact, exchanging questions and responses to facilitate a timely review (primarily questions and responses relating to understanding of the data and analyses presented).

For the safety review, the NRC staff will examine DOE's Preliminary Safety Analysis Report (PSAR) to determine whether the repository site and design are safe and consistent with applicable rules and regulations; whether valid methods of evaluation were employed and accurately carried out; and whether DOE conducted its analysis and evaluation in sufficient depth and breadth to support staff approval with respect to safety. When the staff is satisfied that its own acceptance criteria have been met by the PSAR, the staff will prepare a Safety Evaluation Report (SER) which summarizes the results of its review regarding the anticipated effects of the proposed repository on public health and safety. Following publication of the SER, the NRC's Advisory Committee on Reactor Safeguards (or, in the repository case, an equivalent advisory committee established for waste management) will prepare its own independent report and recommendations to the NRC Chairman, and the staff will issue a supplement to the SER incorporating any changes or actions adopted as a result of the ACRS recommendations. As a result of NRC's questions throughout the review and its SER and any supplements thereto, DOE will issue supplements amending its PSAR. When the NRC staff and the advisory committee have completed their reviews, a public hearing on the safety aspects of the decision will be held near the proposed repository site (NRC, 1987). The hearing stage is a significant contributor to the time and data base requirements involved in a licensing proceeding. This stage can go on from several years to many years depending on the complexity of the case and the number and difficulty of the contentions. As the hearing stage progresses, both DOE's Safety Analysis Report and NRC's Safety Evaluation Report will be finalized. After the hearing record is closed, each of the parties will submit to the ASLB proposed findings of fact. The ASLB will consider each of the findings and will file a decision on the case.

NRC's environmental review of DOE's Final Environmental Impact Statement (FEIS) will proceed in parallel with the safety review and will focus on environmental and site suitability aspects of the proposed facility. The NWPA, as amended, requires that NRC adopt, to the extent practicable, DOE's environmental impact statement. Unlike the typical reactor licensing case, DOE will have issued a Draft Environmental Impact Statement one and a half years before license application is submitted to NRC. This allows more time for review and comment by Federal, state and local agencies, other interested parties and members of the public, as well as for the conduct of public hearings. Following NRC's review of the FEIS, the ASLB will conduct a hearing on the environmental impact and site suitability aspects, following similar steps as outlined above.

If the staff reviews and ASLB hearings have resulted in favorable findings, an authorization to construct the repository could be granted. The current DOE estimate for this milestone is early 1998. Once an NRC authorization to construct the repository is granted, DOE may begin repository construction.

Before DOE can accept spent fuel for disposal in the repository, it must submit an updated license application to the Commission. Since the repository will operate in two phases (Phase I for disposal of limited quantities of spent fuel and Phase II for full operation), DOE will have to submit an updated license application twice. The same process described above will take place; however, the timespans involved are expected to be shorter. The first amended application could be submitted in the year 2000, with receipt of a license to receive and possess limited quantities of radioactive material in 2003. The second amended application could be submitted in 2003, with receipt of a license to receive and possess additional quantities of radioactive material in 2006. (DOE, 1987c). However, these dates are not yet firmly decided.

All these schedules are dependent on the outcome of challenges, motions and information submitted during the NRC licensing proceedings.

3.1.3 Summary of Documents to be Generated In Support of the High-Level Waste Repository Licensing Proceeding

A significant portion of documents in support of the high-level waste licensing proceeding will be generated during the site characterization phase. While many of the major types of documents can be identified at this time, DOE and NRC are still in the planning stages with respect to the timing, nature and flow of documents during this phase. Figure 3 illustrates one of the preliminary concepts being considered within DOE for the flow of documents during site characterization. Based on the issues identified in the DOE issues hierarchy, the SCP identifies the information needed to resolve each of the issues. After more detailed study plans and technical procedures have been developed by DOE and approved by NRC, detailed site characterization activities (e.g., site investigations, advanced design activities, and performance assessment) will be undertaken and will continue for the next four years to support development of the license application.

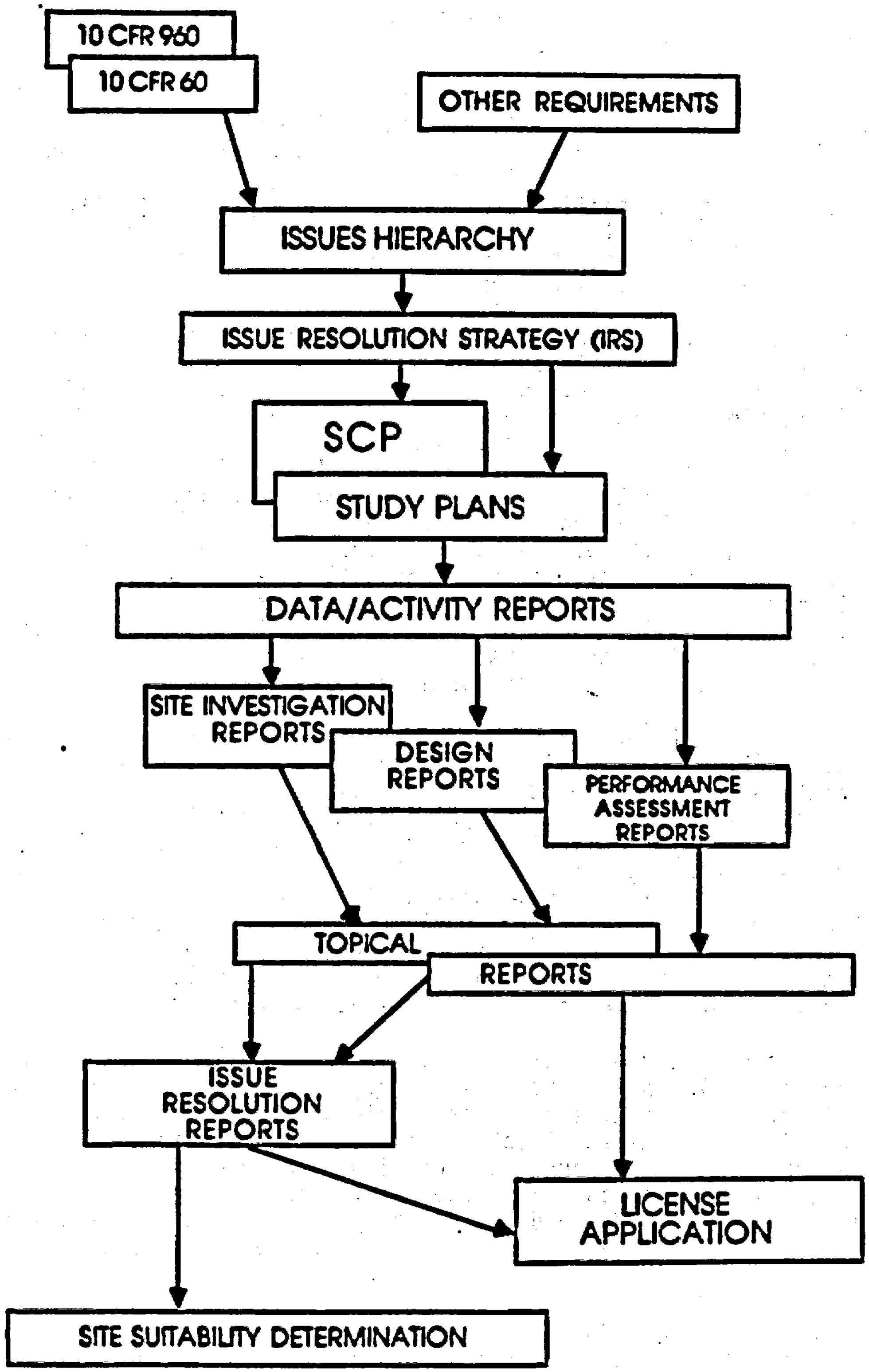
Raw data will be collected and screened for each technical activity identified in the SCP and the data will be summarized and analyzed in data reports (also referred to as activity reports or data/activity reports). Chapter 8 of the CDSCP identifies each of the technical activities and milestones. It is likely that some type of data report will be generated at each milestone or activity point. This can result in hundreds to thousands of data reports of varying length and with varying amounts of figures, maps, calculations and drawings.

To integrate the results of the data reports under each area of investigation in the SCP, DOE will then prepare higher levels of reports, which summarize the data and present conclusions relative to the information needs identified in the SCP. This next level of reports will consist of three types:

(1) site investigation reports,
(2) design reports,
(3) performance assessment reports.

The next reporting level will be topical reports, which will integrate and analyze the data and conclusions presented in the site investigation, design, and performance assessment reports for individual regulatory topics; summarize the data and conclusions relative to each topic; and state the DOE licensing position for each topic. These reports are analogous to "position papers" and will serve as a vehicle through which DOE establishes and communicates its developing technical information base for demonstrating compliance with regulatory requirements. These reports will also serve as the basis for interactions between DOE and NRC staff, allowing early NRC comment on the developing positions. Finally they will provide information to interested parties and will contribute to the development of the next level of report (DOE, 1988a).

Figure 3.
A PRELIMINARY CONCEPT OF DOCUMENT FLOW DURING SITE CHARACTERIZATION



Issue resolution reports will utilize the licensing positions presented in approved topical reports and the data/information from their supporting technical reports to demonstrate the resolution of the "issues" and "key issues" of the DOE issues hierarchy.

In sum, the reports identified above provide "building blocks" for the issue resolution process. An individual issue resolution report could potentially contain input from a number of topical reports, and any one topical report could be applicable in several different Issue Resolution Reports. The same applies to the lower-level documents. The issue resolution reports will be used in confirming site suitability and will provide supporting documentation to the license application. Again, the program is evolving and plans are not yet firm.

The reports described above are the primary issue resolution documents, which will support the license application, including the PSAR and FEIS. However, the universe of documents to be generated during the site characterization phase is even larger. The reports generated by DOE will spawn other documents generated by NRC, the potential host State, other interested parties, and perhaps, again, DOE. In addition, the other involved parties will be generating their own independent reports, which may follow a similar "building block" approach. Also to be accounted for are the documented results of the numerous technical and management meetings, workshops and other interactions that will take place during this period. Table 2 provides a more detailed listing of the specific types of documents or records that will be generated during the site characterization phase, and Figure 4 provides an estimated timeline for document generation during site characterization and up to submittal of a license application.

The volume of documents generated is expected to be the heaviest from 1990 to 1995, when DOE is generating site characterization data and incorporating the interpreted data and analyses into the license application (including the DEIS, FEIS and PSAR). If there is not a significant number of legal challenges following submission of the license application, the rate of new document generation would be expected to decrease. However, it is probable that legal challenges will intensify during the three-year licensing period (e.g., 1995 to 1998). The challenges themselves could lead to the need to generate more data and documents. As a prudent measure, one should assume that the generation of new documents will either remain at the same level as during the site characterization phase or increase.

The scope of such documents will include the legal documents normally involved in a licensing proceeding (e.g., contentions, interrogatories, depositions, testimony, affidavits, motions, findings of fact, decisions), the technical documents involved in the NRC review of the license application (e.g., many rounds of questions and responses, amendments to the Environmental Impact Statement, the Safety Analysis Report, and the Safety Evaluation Report), and the difficult-to-predict technical data and reports that will have to be generated as a result of challenges to the DOE proposal.

Neither DOE nor NRC have reached a stage where it can identify the precise timing of the generation of these licensing documents or their volume.

TABLE 2 - PRELIMINARY LIST OF THE MAJOR TYPES OF DOCUMENTS TO BE GENERATED BY THE PROGRAM

(Repository/MRS/Transportation)

MAJOR PROGRAM DOCUMENTS

Site Survey & Area Recommendation Reports
Environmental Assessments (Draft/Final) and comments
Site Characterization Plan (Draft/Final) and comments
Site Characterization Semi-Annual Reports
Repository, MRS and Waste Package Designs
Spent Fuel Shipment Cask Design
Safety Analysis Report Packages (Shipment Casks)
Environmental Impact Assessments (Draft/Final) and comments
Recommendation Report to President
Notices of Disapproval by State
License Application
- Safety Analysis Report and updates (DOE)
- Safety Evaluation Reports and updates (NRC)

Records of Proceedings (contentions, interrogatories, testimony, evidence, proposed findings of fact, ASLB decision, etc.)
NRC Construction Authorization and Cask Certifications
NRC Authorization to Receive and Possess Radioactive Material

SUPPORTING PROGRAM DOCUMENTS

DOE Study Plans and Technical Procedures
DOE Institutional Plans and Procedures
DOE Data/Activity Reports
DOE Site Investigation Reports
DOE Design Reports
DOE Performance Assessment Reports
DOE Regulatory Topical Reports
DOE Issue Resolution Reports
NRC Comments on above
State reports and comments
Public comments
NRC reports (results of contractor analyses, etc.)
Public hearing records
Contractual Statements of Work
Transportation Routes and Route Survey

REGULATIONS, GUIDANCE AND POLICY

Petitions for Rulemaking
Rulemakings
- Advanced Notices of Proposed Rulemaking
- Public Comments on Advanced Notices
- Proposed Rulemaking
- Public Comments on Proposed Rule
- Final Rule and Agency Analyses of Comments

TABLE 2 - PRELIMINARY LIST OF THE MAJOR TYPES OF DOCUMENTS TO BE GENERATED BY THE PROGRAM

(Repository/MRS/Transportation)
(continued)

NRC Technical Positions
NRC Regulatory Guides
Standard Review Plans
Requests for legal interpretations
Responses to legal interpretations
Congressional testimony (DOE, NRC, State, intervenors)
Responses to Congressional questions (DOE, NRC, State, intervenors)
Testimony before independent review boards/advisory committees
Responses to questions from independent review boards/advisory committees
Reports and recommendations by independent review boards/advisory committees

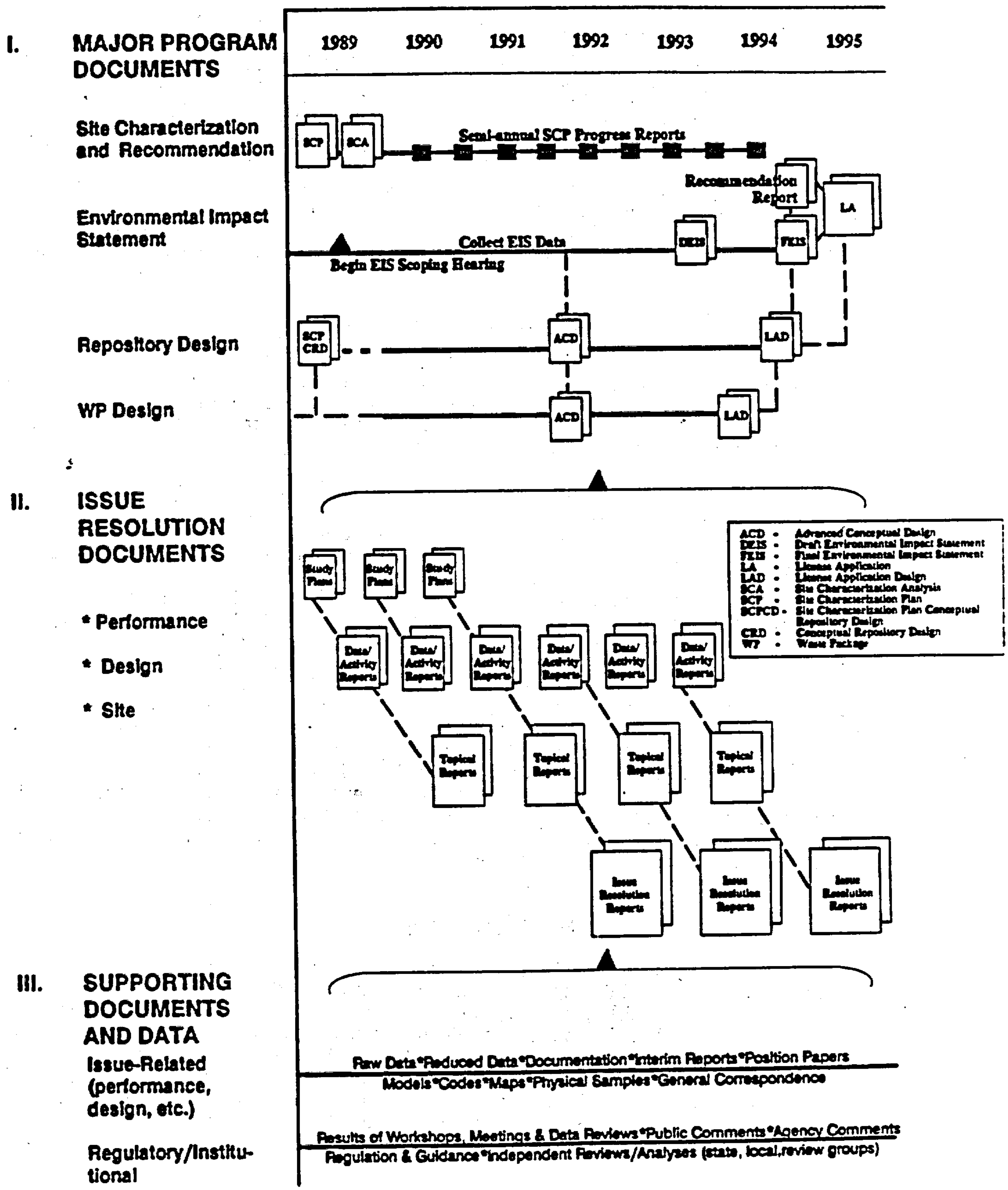
OFFICIAL COMMUNICATIONS BETWEEN AGENCIES

Summaries of NRC/DOE Management Meetings and resulting correspondence
Summaries of NRC/DOE/State Technical Meetings and resulting correspondence
Summaries of NRC/DOE/State Technical Consultations and resulting correspondence
DOE/NRC/State Technical Briefings and resulting correspondence
NRC/State Data Examinations (at site) and resulting correspondence
Summaries of Site Visits and resulting correspondence
QA documentation (including plans/procedures, audit reports, comments by observers, responses to audit reports, etc.)

INFORMAL COMMUNICATIONS BETWEEN AGENCIES

One-on-one correspondence between agency staffs (DOE, NRC, State, etc.)
Summaries of internal NRC meetings
Summaries of internal DOE meetings
Summaries of internal State meetings
Reports by NRC's On-Site Licensing Representative

Figure 4.
SUMMARY OF PROPOSED DOCUMENT GENERATION DURING SITE CHARACTERIZATION



3.2 Current Data Bases of Importance to LSS

There currently exists three large collections of data that meet in part the information needs of some potential LSS users and therefore should be included in the LSS data base. Data held by OCRWM headquarters is indispensable to potential LSS users because it includes documentation of OCRWM policy decisions. The OCRWM collection in Las Vegas is essential to potential LSS users because of the Nevada-specific focus of its technical document collection. The substantial collection of nuclear waste management documents at NRC is also important to potential LSS users. The richness of this collection provides a valuable supplement to the OCRWM document collections. The following sections briefly describe these holdings and indicate the contribution they would ultimately make to the LSS collection.

3.2.1 Profile of the OCRWM Headquarters Data Base

In February 1987 work began on processing OCRWM headquarters documents and records. A bibliographic data base, called the Automated Records System (ARS) was established, containing description of documents contained on microfilm. ARS is used to support litigation and provide on-going headquarters records management support.

Table 3 summarizes the types of documents described in the ARS. Table 4 shows several statistics about the document types. Approximately 657,000 pages are contained in the document collection for which bibliographic descriptions are available in ARS. There currently exists a backlog of about 162,000 documents in addition to these documents. Using 5.8 pages per document as an average (derived from a sample of the data base) an estimate of 945,000 pages of backlog can be estimated. OCRWM staff is currently in the process of screening criteria for document entry into the LSS. An initial scan of the data base has identified some extraneous materials including various communications with the public, annual reports, cost reports, administrative reports, (and internal reviews that are not pertinent to the LSS. Also, among the documents that are pertinent to LSS there is substantial duplication. In many instances there are more than a dozen copies of the same document stored in different OCRWM staff offices. The portion of the data base that is relevant and not duplicative is important in that it provides documentation of OCRWM policies and decisions. As Table 4 reveals, more than 60% of the collection is in the form of letters, memos, telefaxes, and other correspondences.

3.2.2 Profile of the NNWSI Data Base

The Las Vegas collection is clearly much more scientific and technical in nature than the OCRWM headquarters collection. Many of these reports are contained in the backlog of 845,000 documents. These reports are thought to be longer than the 5.8 pages per document average at OCRWM headquarters. Estimates of their size range between 8 and 10 pages per document. These estimates produce a backlog at Las Vegas of 6.8 to 8.5 million pages. A much larger portion of the documents in this collection is considered to be potentially useful to LSS users than in the OCRWM Headquarters ARS. With duplicates eliminated, 65% to 70% of the documents in this collection appear

TABLE 3. TYPES OF DOCUMENTS CONTAINED IN THE OCRWM HEADQUARTERS ARS DATA BASE

ARS Document Type	Description
Correspondence	Includes both incoming and outgoing communications such as letters, memos, and telefaxes.
Reports	Summaries of administrative and technical subject. Can include results of technical investigations, status reports, speeches, hearings.
Publications	Commercially published non-graphic material such as books, journals, Articles, Federal Register and newspaper.
Governing Documents	Includes directives, procedures, orders, resolutions, regulations, instructions, plans, policies, and guidelines. (Not reports discussing directives, plans, etc.)
Graphics	Generally consists of non-textual materials such as maps, charts, drawings, photographs.
Procurement	Documents related to the acquisition, inspection, maintenance, and funding of program activities. This includes RFP's, proposals, contracts, bidder lists, supplier/vendor lists.
Raw Data	Unanalyzed data/information primarily of a technical nature which may appear on logs or stripcharts.
Legal	Includes legal documents produced in the litigation process, patents, and agreements.

TABLE 4. CONTENTS OF THE OCRWM HEADQUARTERS AUTOMATED RECORD SYSTEM DATA BASE, MARCH 10, 1988

ARS Document Type	Number of Documents	% of Total	Average Page Length	Pages
Correspondence	69,971	62	4.8	337,300
Reports	29,386	26	5.4	157,500
Publications	6,480	6	4.6	29,500
Governing Documents	3,022	3	13.9	42,000
Graphics	2,815	2	7.6	21,300
Procurement	808	1	8.1	65,100
Raw Data	420	0	2.9	1,200
Legal	246	0	13.6	3,400
TOTAL	113,088	100%	5.8	657,300

to be appropriate for inclusion in the LSS data base. OCRWM will shortly begin a detailed investigation of this data base to verify Las Vegas estimates.

3.2.3 Profile of the Nuclear Regulatory Commission Data Base

The Transitional Licensing Support System at the Nuclear Regulatory Commission contains both the bibliographic records and the full-text of correspondence documents. The documents were primarily prepared by NRC's Division of Waste Management and their contractors. To date, 3,000 of these documents are included in the system. An additional 47,000 documents are in the backlog. The documents average 7 pages per document, thirty documents per day are being acquired for entry in the system.

3.3 Contents of the LSS Data Base

In this section the characteristics and scope of information required by potential users of LSS are described. Much of this description is derived from the recent Preliminary Needs Analysis that was based on interviews conducted in early 1988. This description is followed by an examination of some key features of the LSS data base.

3.3.1 Information Needs of Potential Users

Four LSS end-user groups are described in the Preliminary Needs Analysis (DOE, 1988d). Each of these groups comprise usage patterns reflecting similar traits. These end-usage categories are:

- Technical and Engineering Usage
- Regulatory and Licensing Usage
- Management and Administrative Usage
- Public Information and General Public Usage.

In addition to the four end-user groups, two other groups were identified:

- Intermediary Usage
- Data Base Management and Quality Assurance Usage.

The characteristics of and data requirements for these end-user groups are provided below.

3.3.1.1 Technical and Engineering Usage

This is the largest usage group constituting 45% of total usage according to the Preliminary Needs Analysis. Scientists and engineers are characteristic of this usage type. Many of these users will require information during the preparation and review of technical reports used in support of the licensing process. Others, more removed from the formal process, will wish to access the LSS with no less professional interest. Members of this usage group will be primarily from the technical staff of federal agencies, national laboratories, state and local agencies, environmental and public interest groups, and the contractors supporting these sub-groups. Their questions will deal mostly with primary data, published analyses of technical issues, computer program documentation, quality assurance procedures and testing procedures. The documents sought by these scientists and engineers will be DOE, NRC and national laboratory technical reports, articles in scientific and engineering journals, progress and summary reports of contracts for government agencies.

The analytical, experimental, and scientific orientation of this group will typically be concentrated on specific topics and issues. They will use bibliographic citations as a means for conducting comprehensive searches of their topical data bases. They will be concerned with the supporting evidence for claims made in their literature. The chain of citations which they need to examine will vary, but will certainly emanate from or lead to the key documents to be generated by the OCRWM program (described in Section 3.1 above). The documents referenced in the key documents should themselves be included in the LSS data base, because of the substantial likelihood that they will be consulted by these users. Without their inclusion in LSS, research time will be increased. Also, acquisition of many of the relevant technical reports is so difficult that some may not be obtained by researchers. Failure to include the key LSS documents in the data base may retard the licensing process and decrease the quality of scientific and engineering support and review of the license and construction authorization applications.

The argument for inclusion of second generation citations in the LSS data base (documents that are referenced in the documents referenced in the key licensing documents) has some merit but is less persuasive. The contention that technical and engineering documents pertinent to OCRWM issues should be included along with the supporting background documents from the literature of the relevant disciplines (e.g., geology, geoengineering, hydrology, geochemistry) is more easily defended.

This analysis of what should be included in the LSS is built upon the earlier needs analysis. These conclusions should not be interpreted as limiting inclusion of technical and engineering documents to less than the needs of all potential users. These conclusions are an effort to translate the global expressions of need into an operational set of data scope requirements.

3.3.1.2 Regulatory and Licensing Usage

Usage in this category, about 25% of all usage according to the Preliminary Needs Analysis,is expected to be primarily by regulatory and licensing specialists (including legal staff) requiring access to both technical and regulatory information. These users are procedure- and strategy-oriented, with a broad qualitative bent. Their concerns are with defensability of positions, completeness of documentation, and direction of overall policies and strategies.

Before submittal of the license application, this group has to perform three major regulatory functions. First, regulatory support staff will perform an on-going oversight role to ensure that technical work will result in a complete and defensible license application. Second, the regulatory support staff will direct and participate in topical report development, seeking early resolution of issues. Third, programmatic decisions must be reviewed by legislative/policy analysts to determine if actions contemplated are within the letter and intent of applicable federal, state, and local laws and regulations.

After submittal of the license application, the licensing support staff will be responsible for developing positions on hearing issues, identifying witnesses, preparing testimony, responding to motions, etc.

Some of the documents sought by users in this usage group include technical and non-technical reports, correspondence, meeting minutes, regulations, regulatory guidance, planning documents, and commitments. Of the LSS subsystems, the Regulations Access Subsystem would be the most often used by the regulatory and licensing usage group. However, heavy usage of the Records Access Subsystem and Issues and Commitment Tracking subsystems will also be made.

3.3.1.3 Management and Administrative Usage

Usage in this category (about 5% of all usage according to the Preliminary Needs Analysis) is expected to be primarily by managers and administrators who are concerned with projects and contracts they are conducting or monitoring. This group will be mainly task, project and program managers and administrators as well as line managers from government agencies, national laboratories, and private contractors. Some of the documents sought by these users include planning documents, cost and schedule performance charts, statements of work, quality assurance audit reports, correspondence, action and commitment tracking documents and memoranda of understanding.

A data base useful to users of this type currently exists in the Automated Record Systems in Washington and Las Vegas. The information system for LSS must provide broader and larger user community access and retrieval features and will achieve this, in part, through the development of the Issues and Commitments Tracking Subsystem. Nevertheless, the scope of the data base required to satisfy this usage group is of the kind already available. That data base will expand as licensing activities progress but its general characteristics are likely to be similar to the current ARS data base.

3.3.1.4 Public Information and General Public Usage

Usage in this category (about 5% of all usage, according to the Preliminary Needs Analysis) is expected to be primarily in support of information needs of the general public, either in response to direct inquiry or through information dissemination by public information specialists. These users' questions will deal mostly with general and descriptive information about nuclear waste management and OCRWM activities, and summary information on technical and environmental issues. Some of the documents sought in this usage category are records of public hearings, issue papers, summary technical documents, fact sheets, documents open for public comment, periodical articles and press releases.

Usage by this category can be expected to expand as the licensing process proceeds. Controversy on the repository is likely to develop and generate a substantial data base that needs to be incorporated in the LSS. Newspaper and magazine articles, books, press releases, fact sheets, issue papers will all proliferate in response to heightened public interest. Provision should be made in the LSS data base to include these documents and to assure complete representation of the various points of view that will emerge.

3.3.1.5 Other Usage Groups

During the development of the LSS Preliminary Needs Analysis, it became apparent that, in addition to the four potential LSS end-user groups discussed above, there were two other groups of potential LSS users:

(1) An Intermediary Usage Group
(2) A Data Base Management and Quality Assurance Usage Group.

These two groups jointly constitute the remaining 20% of LSS usage according to the Preliminary Needs Analysis.

3.3.1.5.1 Usage by Intermediaries

Usage in this category is expected to be primarily in support of information needs of the end-usage categories, generally in response to inquiries by those who do not have direct access to LSS or those who do not want to access the system themselves. This intermediary group will be mainly librarians and information specialists and may also include administrative assistants, researchers and paralegal staff who have a working knowledge of information retrieval or have become thoroughly experienced in searching the LSS.

The documents sought in this usage category will be all of the documents in the system needed by users in each of the four end-usage categories. While no additional documents will be required by these intermediaries, their greater sophistication in information system usage can be expected to impact on the supporting data systems. These users are more likely to exploit the full capabilities of structured indexed searching and to press the limits of full-text searching.

3.3.1.5.2 Data Base Management and Quality Assurance Usage

Usage in this category is expected to be primarily one of controlling and facilitating the flow and quality of data and documents into and out of the LSS. This group will be mainly QA/QC staff, data base maintenance staff and use trainers.

While this group will not require documents in addition to those needed by the end-users, members of this group will need to be supported by a software and QA data base essential for the operation and maintenance of the system.

3.3.1.6 Discovery, Relevance and Privilege

The issues of discovery, relevance, and privilege are important to this preliminary data scope analysis because their interpretation and application to the licensing process are likely to have a significant impact on the types and scope of data to be included in the LSS. Material subject to discovery during repository licensing should be able to be identified through the LSS. Such material must be relevant to the licensing process and must not be subject to any privilege that would limit access.

3.3.1.7 Summary of the Data Needs of Potential Users

In the discussion of each usage group, a general data needs statement has been provided along with examples of specific kinds of documents needed. The number of these users and their needs are likely to expand over time, especially as the licensing process proceeds and public attention becomes focused on repository issues. As scientists and engineers accelerate their research efforts they will probably discover new data needs. Regulatory and licensing activities will also expand and with these an extended array of data needs. And as noted above, increased public focus on the repository will stimulate public information and general public usage.

Under these conditions, it is essential that LSS contain all the data needs of all anticipated users. To plan for less than this is to risk an incomplete data base and extend the licensing period beyond the target of less than three years.

Criteria for the inclusion of information within the LSS must be developed so that an unambiguous decision can be made as to whether a specific document is to be added to the LSS data base. Where the document provides substantial information clearly relevant to the repository licensing or development process the decision to include it is obvious. But if the document provides background information that might be tangentially useful to LSS users, its inclusion is in doubt. It is premature in this preliminary analysis to propose operational rules for document inclusion in the LSS data base. Decisions yet to be made by the NRAC will shape these operational rules. Also, additional analyses of the ARS document collections have begun and these are likely to provide much relevant information. Several items contained in the appendices to this report will be useful to these analytic efforts.

Appendix A includes a page count of the various sections of the Consultation Draft of the Site Characterization Plan (DOE, 1988d). This list may be used as a preliminary indicator of the relative contribution of scientific disciplines to site characterization activities and also as a checklist of some major topics that should be represented in the data base. Appendices B, C, and D provide specific lists of types of documents to be included or excluded from the data base. These lists were prepared by the State of Nevada (Appendix B), the Environmental Defense Fund (Appendix C), and the Department of Energy (Appendix D).

3.3.2 The LSS Regulations Data Base

The Statement of Work for the Licensing Support System (DOE, 1987a) describes the LSS Regulations Access Subsystem as containing the "text of documents that impose legal requirements on the construction or operation of the geologic repository". This definition does not restrict the scope of regulations to the NRC licensing process, and therefore extends to state and environmental regulations that apply to repository siting, construction and operation. In order to estimate the size of the regulatory data base, several documents, which have been prepared for the purpose of defining the regulatory environment of the Yucca Mountain site, were collected and reviewed. These documents include:

1. NNWSI Project Regulatory Document Manual (DOE, 1986a)
2. Consultation Draft Site Characterization Plan (DOE, 1988b)
3. Environmental Assessment Yucca Mountain Site (DOE, 1986b)
4. Generic Requirements for Mined Geologic Disposal System (Weston, 1986)
5. Draft Environmental Regulatory Compliance Plan for Site Characterization of the Yucca Mountain Site (DOE, 1988a).

In addition to these sources, discussions with NRC and SAIC staff members familiar with the regulatory environment yielded additional insight into the extent of the applicable regulations and statutes.

The result of this effort is shown in Table 5, a composite listing of potentially applicable regulations. The major sources of regulations include:

1. The Code of Federal Regulations (CFR)
2. The United States Code (USC)
3. Nevada Administrative Code
4. Nevada Revised Statutes
5. Executive Orders.

In addition, certain documents that provide guidance are included such as NRC generic technical positions and DOE orders. The total number of pages included in the regulations listed in Table 6 is 8,311. One should be cautioned in extrapolating the page count to a character count as the number of characters per page varies widely depending on the format. NRC technical positions, for example, average about 2500 characters per page while regulations appearing in the small print, three-column format of the Federal Register average close to 7500 characters per page.

In estimating the size of the regulations data base, it is necessary to define a model for the incorporation of the regulations into the LSS. The model used in this analysis is based on the premise that the initial LSS data base will include all of the pertinent regulations as published in the most recent issue of the appropriate document (Code of Federal Regulations,

TABLE 5. REGULATIONS RELEVANT TO YUCCA MOUNTAIN SITE REPOSITORY PROGRAM

Notes:

1) "Where cited" references:

   N - NNWSI Project Regulatory Document Manual
   S - Draft Consultation Site Characterization Plan
   E - Environmental Assessment
   R - Generic Requirements for a Mined Geologic Disposal System
   C - Draft Environmental Regulatory Compliance Plan for Site Characterization of Yucca Mtn Site (NNWSI)
   O - Other sources (Informal discussion and interviews)

2) The letter "e" following a page count indicates an estimate.

CITATION	TITLE	WHERE CITED N	S	E	R	C	O	PAGES
	7 CFR - AGRICULTURE							
7CFR658	Farmland Protection Policy Act					X		6
	10 CFR - ENERGY							
10CFR2	Rules of Practice for Domestic Licensing Proceedings						X	118
10CFR20	Standards for Protection Against Radiation	X	X	X	X		X	26
10CFR21	Reporting of Defects for Noncompliance	X						3
10CFR50App A	General Design Criteria for Nuclear Power Plants		X					9
10CFR50App B	Quality Assurance Criteria for Nuclear Power Plants and Fuel Reprocessing Plants	X	X		X			4
10CFR50App I	Numerical Guides for Design Objectives and Limiting Conditions for Operation to Meet the Criterion "As Low as is Reasonably Achievable" for Radioactive Material ...		X					4
10CFR51	Environmental Protection Regulations for Domestic Licensing and Related Regulatory Functions	X					X	19
10CFR60	Disposal of High-Level Radioactive Waste in Geologic Repositories	X	X	X	X		X	31
10CFR71	Packaging and Transportation of Radioactive Material	X		X			X	30
10CFR72	Licensing Requirements for the Storage of Spent Fuel in an ISFSI		X				X	24
10CFR73	Physical Protection of Plants and Materials			X	X		X	58
10CFR100	Reactor Site Criteria		X					4
10CFR100AppA	Seismic and Geologic Siting Criteria for Nuclear Power Plants		X					8

TABLE 5. REGULATIONS RELEVANT TO YUCCA MOUNTAIN SITE REPOSITORY PROGRAM (continued)

CITATION	TITLE	WHERE CITED N	S	E	R	C	O	PAGES
10CFR960	General Guidelines for the Recommendation of Sites for Nuclear Waste Repositories; Final Siting Guidelines	X	X	X	X		X	56
10CFR961	Standard Contract for Disposal of Spent Nuclear Fuel and/or High Level Waste		X	X				18
10CFR1021	Compliance with the National Environmental Policy Act						X	1
10CFR1022	Compliance with Federal Floodplain/Wetlands Environmental Review Requirements					X		4
								----- 415
	25 CFR - INDIANS							
25CFR261	Preservation of Antiquities					X		2
	29 CFR - LABOR							
29CFR1910	Occupational Safety and Health Standards				X			918
29CFR1926	Safety and Health Regulations for Construction				X		X	292
								----- 1210
	30 CFR - MINERAL RESOURCES							
30CFR ChapI Parts5-100	Mine Safety and Health Administration, Department of Labor		X		X		X	668
	33 CFR - NAVIGATION & NAVIGABLE WATERS							
33CFR209	Administrative Procedure					X		52
33CFR320	General Regulatory Policies					X		13
33CFR323	Permits for Discharges of Dredged and Fill Material into Waters of the U.S.				X	X		6
33CFR324	Permits for Ocean Dumping of Dredged Material					X		2
33CFR325	Processing of Dept of the Army Permits					X		18
33CFR326	Enforcement					X		4
33CFR327	Public Hearings					X		3
33CFR328	Definition of Waters of the United States					X		2
33CFR329	Definition of Navigable Waters of the U.S.					X		5
33CFR330	Nationwide Permits					X		10
								----- 114

TABLE 5. REGULATIONS RELEVANT TO YUCCA MOUNTAIN SITE REPOSITORY PROGRAM (continued)

CITATION	TITLE	WHERE CITED N	S	E	R	C	O	PAGES
	36 CFR - PARKS, FORESTS, AND PUBLIC PROPERTY							
36CFR60	National Register of Historic Places					X		16
36CFR61	Procedures for Approved State/Local Government Historic Preservation Programs					X		15
36CFR63	Determination of Eligibility for Inclusion in National Register of Historic Places					X		3
36CFR65	National Historic Landmarks Program					X		11
36CFR67	Historic Preservation Certification...					X		16
36CFR68	Sec of Interior's Standards for Historic Preservation Projects					X		3
36CFR296	Protection of Archaeological Resources: Uniform Regulations					X		12
36CFR800	Protection of Historic & Cultural Properties					X		17
								90
	40 CFR - PROTECTION OF ENVIRONMENT							
40CFR50	National Primary and Secondary Ambient Air Quality Standards			X	X	X		51
40CFR51	Reqmts for Preparation, Adoption, & Submittal of Implementation Plans					X		92
40CFR52	Approval and Promulgation of Implementation Plans - Subparts A,DD					X		72
40CFR53	Ambient Air Monitoring Ref. & Equiv. Methods					X		47
40CFR58	Ambient Air Quality Surveillance		X	X		X		17
40CFR60	Standards for Performance for new Stationary Sources				X	X		645
40CFR61	National Emission Standards for Hazardous Air Pollutants			X		X	X	71
40CFR81	Designation of Areas for Air Quality Planning Purposes					X		8
40CFR110	Discharge of Oil					X		4
40CFR112	Oil Pollution Prevention					X		12
40CFR116	Designation of Hazardous Substances					X		10
40CFR117	Determination of Reportable Quantities for Hazardous Substances					X		9
40CFR121	State Certification of Activities Requiring Federal License or Permit					X		5
40CFR122	EPA Administered Permit Programs: National Pollutant Discharge Elimination System				X	X		52
40CFR123	State Program Requirements					X		20
40CFR124	Procedures for Decisionmaking					X		54
40CFR125	Criteria and Standards for National Pollutant Discharge Elimination System				X	X		37

TABLE 5. REGULATIONS RELEVANT TO YUCCA MOUNTAIN SITE REPOSITORY PROGRAM (continued)

CITATION	TITLE	WHERE CITED N	S	E	R	C	O	PAGES
40CFR129	Toxic Pollutant Effluent Standards					X		12
40CFR131	Water Quality Standards						X	6
40CFR133	Secondary Treatment Regulation					X		4
40CFR136	Guidelines Establishing Test Procedures for the Analysis of Pollutants					X		273
40CFR141	National Interim Primary Drinking Water Regs.		X	X	X	X		27
40CFR142	National Primary Drinking Water Regulations					X	X	19
40CFR143	National Secondary Drinking Water Regulations				X	X		3
40CFR144	Underground Injection Control Program					X		54
40CFR145	State UIC Program Requirements					X		11
40CFR146	Underground Injection Control Program: Criteria and Standards					X		20
40CFR147	State Underground Injection Control Programs					X		71
40CFR149	Sole Source Aquifers					X		5
40CFR162	Regulations for the Enforcement of the Federal Insecticide, Fungicide and Rodenticide Act				X			68
40CFR190	Environmental Radiation Protection Standards for Nuclear Power Operations			X				2
40CFR191	Environmental Standards for the Management & Disposal of Spent Nuclear Fuel, High-Level and Transuranic Radioactive Wastes	X	X	X	X		X	9
40CFR192	Health and Environmental Protection Standards for Uranium and Thorium Mill Tailings			X				7
40CFR201	Noise Emission Standards for Transportation Equipment			X				13
40CFR204	Noise Emission Standards for Construction Equip				X			17
40CFR220-30	Subchapter H - Ocean Dumping				X	X		74
40CFR240-47	Subchapter I - Solid Wastes				X	X		45
40CFR260-64	Subchapter I - Solid Wastes					X		234
40CFR266	Standards for Management of Specific Hazardous Wastes & Specific Types of Hazardous Waste Management Facilities					X		8
40CFR270	EPA Administered Permit Program: Hazardous Waste Permit Program					X		42
40CFR271	Requirements for Authorization of State Hazardous Waste Programs					X		34
40CFR280	Underground Storage Tanks					X		7
40CFR300	National Oil and Hazardous Substances Pollution Contingency Plan					X	X	96
40CFR302	Designation, Reportable Quantities, and Notification					X	X	80
40CFR355	Emergency Planning and Notification					X		24
40CFR401	General Provisions					X		5
40CFR403	General Pretreatment Regulations for Existing and New Sources of Pollution					X		36
40CFR1500-08	Subchapter V - Council on Environmental Quality				X	X	X	29

								2534

TABLE 5. REGULATIONS RELEVANT TO YUCCA MOUNTAIN SITE REPOSITORY PROGRAM (continued)

43 CFR - PUBLIC LANDS: INTERIOR

CITATION	TITLE	WHERE CITED N	S	E	R	C	O	PAGES
43CFR3	Preservation of American Antiquities					X		2
43CFR7	Protection of Archaeological Resources					X		15
43CFR3600	Mineral Materials Disposal; General					X		8
								25

49 CFR - TRANSPORTATION

CITATION	TITLE	WHERE CITED N	S	E	R	C	O	PAGES
49CFR101	Office of Transportation Security - Cargo Security Advisory Standards				X	X		19
49CFR106-07	Subchap B - Hazardous Materials Transportation and Pipeline Safety					X	X	27
49CFR171-78	Subchap C - Hazardous Materials Regulations		X	X		X	X	1189
								1235

50 CFR - WILDLIFE AND FISHERIES

CITATION	TITLE	WHERE CITED N	S	E	R	C	O	PAGES
50CFR17	Endangered and Threatened Wildlife and Plants					X		139
50CFR402	Interagency Cooperation - Endangered Species Act of 1973					X		12
50CFR424	Listing Endangered and Threatened Species and Designating Critical Habitat					X		8
50CFR450-53	Subchap C Endangered Species Exemption Process					X		9
								168

PUBLIC LAWS

CITATION	TITLE	WHERE CITED N	S	E	R	C	O	PAGES
	Nuclear Energy:							
42USC2011-2284	The Atomic Energy Act of 1954 as Amended	X		X			X	137
42USC5801 et seq	Energy Reorganization Act of 1974	X		X			X	21
	Department of Energy Organization Act	X		X				49
42USC10101 et seq	Nuclear Waste Policy Act of 1982	X			X		X	63
	Nuclear Waste Policy Amendments Act of 1987	X					X	29
	Environment:							
15USC2601-2654	Toxic Substance Control Act				X			20 e
16USC470 et seq	National Historic Preservation Act			X		X		27
16USC1531-1542	Endangered Species Act			X		X		33
30USC601-604	Materials Act					X		2
33USC1311-1376	Federal Water Pollution Control Act				X	X	X	57
42USC300f-300j-1	Safe Drinking Water Act				X	X	X	29
49USC1801-1812	Hazardous Material Transportation Act			X		X	X	5
42USC1996	American Indian Religious Freedom Act					X		2 e
42USC4321-4347	National Environmental Policy Act	X			X	X	X	20

TABLE 5. REGULATIONS RELEVANT TO YUCCA MOUNTAIN SITE REPOSITORY PROGRAM (continued)

CITATION	TITLE	WHERE CITED						PAGES
		N	S	E	R	C	O	
========	=====							=====
42USC4901-4915	Noise Control Act, Quiet Communities Act				X	X		10
42USC6901-6979	Hazardous and Solid Waste Amendment Act, Resource Conservation & Recovery Act				X	X	X	42
42USC7401-7428	Clean Air Act		X	X		X		57
42USC9601-9657	Comprehensive Environmental Response Compensation and Liability Act					X	X	32
	Land:							
7USC4201-4209	Farmland Protection Policy Act			X		X		2
43USC1701-1771	Federal Land Policy and Management Act			X		X		33
	Safety:							
30USC801-878	Federal Mine Safety and Health Act						X	46

								713
	STATE REGULATIONS							
	Air Quality:							
NevAdminCode	Sections 445.430-445.995					X		50 e
	Water Pollution/Underground Injection Control:							
NevAdminCode	Sections 445.2-445.96					X		50 e
NevAdminCode	Sections 445.140-445.182					X		20 e
NevAdminCode	Sections 445.224-445.420					X		50 e
NevRevStatutes	Sections 445.131-445.354					X		54
	Appropriate Public Waters:							
NevRevStatutes	Section 533.325					X		1
	Wildlife:							
NevRevStatutes	Sections 501.105-501.110					X		1
NevAdminCode	Sections 503.010-503.080					X		5 e
	Vegetation:							
NevRevStatutes	Sections 504.520, 527.050, 527.100, 527.105 527.260, 527.270, 527.500					X		5

								236
	EXECUTIVE ORDERS							
12088	Federal Compliance with Pollution Control Stds				X			3
11988	Protection of Floodplains					X		3 e
11990	Protection of Wetlands					X		3 e

								9
		Total Regs						7423

TABLE 5. REGULATIONS RELEVANT TO YUCCA MOUNTAIN SITE REPOSITORY PROGRAM (continued)

CITATION	TITLE	WHERE CITED N	S	E	R	C	O	PAGES
	GUIDANCE DOCUMENTS DOE ORDERS							
5440.1C	Implementation of NEPA				X			9
5480.1A	Environmental Protection, Safety and Health Protection Programs for DOE Operations				X			50 e
5480.4	Environmental Protection, Safety and Health Protection Standards				X			25 e
5482.1A	Environmental Protection, Safety and Health Protection Appraisal Program				X			11
5484.1	Environmental Protection, Safety and Health Protection Information Requirements Reporting				X			40 e
5484.2	Unusual Occurrence Reporting System				X			20 e
6430.1	General Design Criteria				X			250 e
								405
	NRC DOCUMENTS							
REGUIDE 4.17	Standard Format and Content of Site Characterization Plans for HLW Geologic Repositories	X						48
	NRC Final Generic Technical Positions:							
NUREG-0856	Documentation of Computer Codes	X						12
	Radionuclide Solubility in Groundwater	X						17
	In Situ Testing during Site Characterization	X						18
	Design Information Needs in the Site Characterization Plan	X						13
	Waste Package Reliability Analysis	X						26
	Determination of Radionuclide Sorption	X						19
	Borehole and Shaft Sealing	X						20
	Revised Modeling Strategy Document for HLW Performance Assessment	X						65
	Qualification of Existing Data	X						4
	Peer Review	X						5
	Draft Generic Technical Positions (in Total)	X						236
								483
		Grand Total						8311

U.S. Code, etc.). The LSS will not separately contain background information leading up to those regulations. For example, changes in the CFR are published in the Federal Register along with discussions covering comments, background, comment resolution, and other information pertaining to the understanding of the interpretation of the regulations appearing in the CFR. Similarly, changes in the U.S. Code are based on Public Laws which are passed by Congress. The Public Law itself will not appear in the LSS, unless it has not yet been incorporated into the Code. Following the initial LSS regulations data base loading, the LSS will incorporate all Federal Register notices pertaining to the listed regulations or deemed relevant, all Public Laws affecting the U.S. Code citations in the data base or deemed relevant, and the Regulations Access system data base will be revised to incorporate all later versions of the documents as they are published.

4.0 ESTIMATING THE SIZE OF THE LSS DATA BASE, 1990-2009

In this section, estimates are made of the amount and nature of the information needed by LSS users in August 1990 when LSS is partially loaded and available. A host of assumptions have been made in developing these estimates. Some have already been stated or are implicit in the preceding discussion. Additional assumptions will be made as the estimates are developed. It is necessary to prepare these estimates in order to design the LSS. However, many of the assumptions made here may be modified by decisions of the NRC Negotiated Rulemaking Advisory Committee or by actions of DOE or NRC.

4.1 Size of the LSS Data Base, August 1990

The data base of the information system being developed by NNWSI is considered in this Preliminary Data Scope Analysis as the core of scientific and technical material required to be contained in the LSS. This is justified because of the completeness of this collection. The estimates of its required size and content (SAIC, 1986b) are used here as the first step in estimating the size of the LSS data base in 1990.

The NNWSI project in 1986 consisted of the following participants:

U.S. Geological Survey
Lawrence Livermore National Laboratory
Los Alamos National Laboratory
Sandia National Laboratories
Science Applications International Corporation
Holmes and Narver
Fenix and Scisson
Reynolds Electrical and Engineering Company
Westinghouse/Waste Technology Services Division.

These project participants and NTS contractors were surveyed to estimate all information relating to the NNWSI project generated by or submitted to these organizations. The survey asked for estimates of the amount of project information that was and would be produced in the period 1980 to 1990. The report (SAIC, 1986b) shows for each participant or contractor the project activity, the document type, and either document count or page count or both. In some cases only the number of pages or documents were provided and the missing quantity was estimated. SAIC staff noted wide differences in the average page per document figures computed from the submitted survey data. Submissions ranged from two pages per document to eight pages per document. Staff increased the estimate to compensate for what they believed was under-reporting in the page count. After discussions with the Las Vegas staff, it was decided to leave the original estimates as reported, but to use in all future estimates for NNWSI documents the more realistic range of 8 to 10 pages per document. Table 6 summarizes the results of the survey in its original form and shows that the data base would consist of about 4 million pages.

These 4 million pages constitute the first step in developing the estimate of the data base needed by LSS users in August 1990. Because the survey included documents to be prepared through Dec. 1989, an adjustment is required to cover the period January to August 1990. At an annual document production rate of 400,000 per year, production over the seven-month period would be about 233,000 (Item 2, Table 7).

In order to add the relevant ARS documents to the estimate without double-counting, the number of NNWSI participants' documents currently in ARS was counted (15,362) and the proportion on non-participant documents was calculated (0.60). As noted earlier, (Section 3.2.2) 65% to 70% of the non-participant documents in Nevada ARS were judged relevant and non-duplicative. When these proportions are combined with the 8 pages per document average (low estimate) and the 10 pages per document average (high estimate) a low estimate of 121,000 and a high estimate of 163,000 of documents referenced in the March 1988 ARS and relevant to LSS is generated (Item 3, Table 7).

New ARS/LAS Vegas documents processed in the 29 months from March 1988 to August 1990 are estimated by using the current ARS monthly production rate of 1,650 documents per month in combination with the factors described in the previous paragraph. Low and high estimates are generated and recorded as Item 4, Table 7.

The current backlog for Las Vegas ARS is 845,000 documents. The number of these documents included in low and high estimates of backlog documents needed by LSS users is described in Item 5, Table 7.

There are 113,088 documents currently indexed in ARS at OCRWM headquarters in Washington, D.C. Only four percent of these documents duplicate the NNWSI project survey documents described in Items 1 and 2 in Table 7. These must be omitted from the count of LSS documents to avoid double counting. Internal duplication of documents at OCRWM headquarters must also be considered as well as duplication with ARS/Las Vegas. Also, as noted in Section 3.2.1 there are many communications, administrative reports and other documents that are not relevant to LSS. Taking these exclusions into consideration, ARS staff at OCRWM estimate that 20% of the collection is relevant to LSS. This percentage is used in preparing the ARS/Washington estimates shown in Items 6,7, and 8 in Table 7.

TABLE 6. ESTIMATES OF THE NUMBER OF DOCUMENTS AND PAGES APPROPRIATE FOR INCLUSION IN THE NNWSI PROJECT DATA BASE FOR THE PERIOD 1980 TO 1990

ORGANIZATION	DOCUMENT COUNT	PAGE COUNT
U.S. Geological Survey	96,860	739,400
Lawrence Livermore National Laboratory	83,355	666,840
Los Alamos National Laboratory	83,780	670,200
Sandia National Laboratories	25,163	201,310
Science Applications International Corp.	27,708	221,667
Holmes & Narver	184,000	368,000
Fenix & Scisson	315,000	630,000
Reynolds Electrical & Engineering	135,870	271,740
Westinghouse/Waste Technology Services Division	87,500	175,000
TOTALS	1,039,236	3,944,157

TABLE 7. SUMMARY OF COMPUTATIONS MADE IN ESTIMATING THE SIZE OF THE LSS DATA BASE IN AUGUST 1990

	Item	Low Estimate	High Estimate
1.	<u>NNWSI Project Documents</u> Data are from 1986 survey of nine organizations that are participants or contractors with NNWSI Project. Pages estimated as produced by them 1980-1990. All documents assumed to be related to only Nevada site.	4,000,000	4,000,000
2.	<u>NNWSI Project Documents for Jan - Aug 1990</u> An estimated 4,000,000 pages for 10 years averages 400,000 pages per year. Estimate for the seven months is 7/12 of 400,000. Assumes production at the mean annual rate over the seven months.	233,000	233,000
3.	<u>ARS/Las Vegas Documents</u> There are 38,868 documents currently described in ARS/Las Vegas. A sampling of ARS showed that 60% were not included in estimates above; and staff estimate that 65% (low estimate) to 70% (high estimate) are relevant to LSS. Estimate of pages per document range from 8 (low estimate) to 10 (high estimate).	121,000	163,000
4.	<u>ARS/Las Vegas Documents, March 1988 - Aug 1990</u> New documents added for the 29 month period at 1,650 documents per month with 60% not duplicated above; and with 65% (low estimate) to 70% (high estimate) relevant. At 8 pages (low estimate) and 10 pages (high estimate) per document.	149,000	201,000
5.	<u>ARS/Las Vegas Backlog</u> There are 845,000 documents in the backlog. Assuming the backlog is similar to documents in ARS, then about 60% of them are not counted in above and 65% (low estimate) to 70% (high estimate) of these are relevant.		

TABLE 7. SUMMARY OF COMPUTATIONS MADE IN ESTIMATING THE SIZE OF THE LSS DATA BASE IN AUGUST 1990 (continued)

	Item	Low Estimate	High Estimate
	Estimate of pages per document are 8 (low estimate) or 10 (high estimate).	2,636,000	3,549,000
6.	<u>ARS/Washington Documents</u> There are 113,088 documents currently indexed in ARS. It is observed that 96% do not duplicate above and 20% of these are judged relevant. A sample of ARS set average pages per document at 5.8.	126,000	126,000
7.	<u>ARS/Washington Documents March 1988 - Aug 1990</u> New documents are added at 50 per day. For the 29 month period at 20 workdays per month, with 96% not duplicating above; with 20% judged relevant and 5.8 pages per document.	323,000	323,000
8.	<u>ARS/Washington Backlog</u> There are 162,000 documents in the backlog. 96% are not duplicative of above and 20% are judged relevant. 5.8 pages per document.	180,000	180,000
9.	<u>NRC's Division of Waste Management Documents</u> The number of documents currently in collection is 50,000. 90% of these were determined as relevant. 7 pages per document.	315,000	315,000
10.	<u>NRC's Division of Waste Management Documents March 1988 - Aug 1990</u> Thirty new documents are being added daily. Assume 20 work days per month for 29 months, 90% relevant, 7 pages per document.	110,000	110,000

TABLE 7. SUMMARY OF COMPUTATIONS MADE IN ESTIMATING THE SIZE OF THE LSS DATA BASE IN AUGUST 1990 (continued)

	Item	Low Estimate	High Estimate
11.	Regulations Data Base Current count of pages is 8,311 (see Section 3.3.2). With revisions and new legislation 9,450 are estimated for August 1990. Only some 10% are duplicated in above estimates.	9,000	9,000
12.	Issues and Commitments Tracking Data Base Estimate of the data base ranges from 4,000 to 5,000 pages.	4,000	5,000
13.	Adjustment for Under-Representation of Relevant Topics Radioactive waste transportation, MRS and socioeconomic effects are insufficiently represented in (1) through (12). An adjustment of 10% of the total through (12) provides a low estimate. An adjustment of 20% provides a high estimate.	821,000	2,304,000
	TOTALS	9,027,000	11,057,000

The number of relevant documents prepared by NRC's Division of Waste Management and its contractors currently total 50,000. These documents average 7 pages per document. Ninety percent of these documents are considered relevant to LSS. These assumptions add 315,000 pages to the sizing estimate of LSS. Again, current production is used to estimate new document production for March 1988 to August 1990 (Items 9, 10, Table 7).

The Regulations data base described in Table 5 is 8,311. Allowing for some growth, 9,450 pages are estimated for August 1990. Only some 10% of these pages are duplicated in above estimates (Item 11).

The size of the required issue tracking and commitments tracking data base as of August 1990 has been estimated by DOE staff to be 4,000 to 5,000 pages (Item 12).

The interim total of pages of documents needed in the LSS data base in August 1990 is based on projections of Nevada specific data combined with (non-redundant) estimates from existing relevant document collections. Detailed analysis on subject content of LSS data base needed by the specific independent end-usage groups have been performed for this Preliminary Data Scope Analysis, by the State of Nevada (Appendix B) and by the Environmental Coalition (Appendix C). The last two were prepared for the March 1988 meeting of the NRC Negotiated Rulemaking Advisory Group. A comparison of the subject contents represented in the interim total with the independently derived subject lists shows that several areas are under-represented in the former. Notable deficiencies in the collective data base are in the areas of radioactive waste transportation, monitored retrievable storage (MRS), and socioeconomic effects. It is also clear that the international literature in general is substantially under-represented as are technical journal articles. These current shortcomings will likely be exacerbated as activity on site characterization, MRS and the transportation of waste picks up pace. Release of the final site characterization plan for Yucca Mountain, scheduled for September 1988, is likely to trigger such activity. Consequently, the documents required to meet these needs should be reflected in the August 1990 LSS data base. Program staff estimate that a 10% to 20% increase in the interim total of documents represents what is needed to fill in current gaps in the coverage of these documents in the estimate and to provide for the rapid growth in document generation in the MRS and transportation areas that will occur between now and August 1990 (Item 13, Table 7).

With the addition of these documents the required size of the LSS data base on August 1990 is estimated to range between 9.8 million and 11.1 million pages. These low and high estimates are based on many assumptions summarized in Table 7. The estimates, while preliminary, give an order of magnitude estimate for design purposes. They will be revised as the results of analyses currently being conducted become available.

4.2 Size of the LSS Data Base, 1990 - 2009

It is anticipated that the per year generation of documents relevant to LSS, will increase and decrease relative to the time-line for key OCRWM activities as shown in Figure 1. Document production will increase in the periods just before the release of important formal deliverables affecting LSS (e.g. the licensing application for the repository, the final environmental impact statements for the repository, and MRS) and will taper off afterwards (i.e. during repository construction).

Table 8 shows low and high projections from year-end 1990 through the following 20-year period. The key assumptions made in preparing this projection are as follows:

(1) Beginning in August 1990 the additional materials appropriate for inclusion in the LSS data base reflects a 10% to 20% increase in annual additions to the data base beyond current levels in order to account for documents in subject areas where coverage in the estimates is inadequate.

(2) The rate of production of documents to be included in the LSS annually varies in accordance with the program activity estimates made in the Preliminary Needs Analysis (notably the relative LSS usage time-line shown in Figure 3c, SAIC, 1988).

The projections show about 21.4 million to 28.0 million pages required by LSS at the end of the century and 31 million to 41 million pages required by the end of the year 2009. The projections are plotted on cumulative distribution curves shown as part of Figure 1.

5.0 CONCLUSIONS

The Conceptual Design Analysis, which is the next in this series of reports, will propose an LSS concept (and variants thereof) consistent with the requirements developed in the Preliminary Needs Analysis and the present Preliminary Data Scope Analysis. The system performance and data base size and content requirements will drive the ultimate LSS design. Several previous studies have attempted to quantify the number of documents (and pages) that constitute the information base available to LSS users. With the passage of time, this task has become easier. A better basis upon which estimates of the quantity and characteristics of the material needed in the LSS now exists, through experience gained in implementing systems such as the NRC Transitional Licensing Support System and the DOE Automated Records Systems, which support users with needs similar to those expected for potential LSS users. Further, the recent legislative action that focuses on a single potential repository site in Nevada simplifies the estimation process and reduces the associated uncertainties.

Generally, the LSS must be able to support all the information needs of all the parties involved in repository licensing (and their staffs and organizations). Specifically focusing on information needs to support licensing, the LSS should provide access to all the key program documents and major milestone documents in the licensing process, along with all of the documents cited in them. Additional supporting material in a variety of relevant technical disciplines should be included to provide an appropriate technical and regulatory context for this core. The liberal interpretation of what should be subject to discovery during the licensing process, which is the conservative interpretation from the perspective of schedule,

TABLE 8. PROJECTION OF THE SIZE OF THE LSS DATA BASE, 1990 - 2009

	LOW ESTIMATE		HIGH ESTIMATE	
Year	Pages Added During Year	Cumulative Pages At Year-End	Pages Added During Year	Cumulative Pages At Year-End
1990	830,000	9,304,000	1,100,000	11,885,000
1991	1,087,000	10,391,000	1,441,000	13,326,000
1992	1,428,000	11,819,000	1,892,000	15,218,000
1993	830,000	12,649,000	1,100,000	16,318,000
1994	2,009,000	14,658,000	2,662,000	18,980,000
1995	1,858,000	16,516,000	2,463,000	21,443,000
1996	1,635,000	18,151,000	2,167,000	23,610,000
1997	1,386,000	19,537,000	1,837,000	25,447,000
1998	623,000	20,160,000	825,000	26,272,000
1999	1,286,000	21,446,000	1,704,000	27,976,000
2000	1,170,000	22,616,000	1,550,000	29,526,000
2001	1,877,000	24,493,000	2,487,000	32,013,000
2002	1,236,000	25,729,000	1,638,000	33,651,000
2003	1,261,000	26,990,000	1,671,000	35,322,000
2004	1,327,000	28,317,000	1,759,000	37,081,000
2005	1,120,000	29,437,000	1,484,000	38,565,000
2006	415,000	29,852,000	550,000	39,115,000
2007	365,000	30,217,000	484,000	39,599,000
2008	365,000	30,582,000	484,000	40,083,000
2009	365,000	30,947,000	484,000	40,567,000

suggests that the criteria for document inclusion should be very broad to insure completeness of the LSS data base. This perspective provides an outer-bound estimate of what could be required.

The major document collections and information systems that have already been developed to support the high level waste (HLW) repository technical and regulatory community provide not only a starting point for estimating the characteristics of the LSS data base, but also contain a significant portion of the documents that currently exist and that will be needed in the LSS collection. Because of the varied needs which these systems were designed to meet, it is clear that not all the information in these systems is relevant for inclusion in the LSS and that the LSS must contain significant material beyond what is in them. A comparison of the subject content of these collections with the requirements expressed for the LSS data base reveals that the collections would not meet the LSS user requirements in a number of areas including primarily transportation, monitored retrievable storage, and socioeconomic effects. This suggests that a concerted effort to strengthen these areas will be required in the process of collecting the backlog material to be included in LSS. The technical material in the current collections also appears to under represent the relevant international literature and articles in peer review technical journals which will be needed by LSS users. Further, it is clear that the requirements of data base completeness for discovery, quality assurance, sophisticated information management system users and the technical breadth of HLW disposal literature lead to the conclusion that an active effort must be made to identify and include information appropriate for storage in the LSS, rather than simply including existing collections. This effort is particularly critical for identifying material that is not yet extensively available, such as quality assurance records and legal material generated during the licensing process.

Based on the concepts and assumptions in this analysis, the estimate of total pages of material that would be needed by potential LSS users (and therefore appropriate for inclusion in the LSS data base) in August 1990 ranges from 9.8 million to 11.1 million pages. Based on further analysis and assumptions, the estimate of material needed increases to between 31 million and 41 million pages over the first 20 years of LSS operation. These estimates are preliminary. While additional analyses of the content and scope of the LSS data base are currently being conducted, and this process will be refined, it is expected to continue for the life of the LSS. As program task schedules and issues become better defined, so will estimates of the scope and quantity of relevant information. However, it may be that the major finding of this study is that the uncertainty of these estimates will continue for some time and that consequently, this uncertainty should be incorporated in the design of the LSS.

REFERENCES

Aerospace, 1986; Requirements Definition for a Licensing Information Management System for Nuclear Waste, Draft Revision 2, The Aerospace Corporation, prepared for NRC, March 1986.

Cotter, B.P., Jr., 1986; Alternative Approaches to Licensing a Geologic Repository, Memorandum to William J. Olmstead, 7 October 1986.

DOE, 1985; Mission Plan for the Civilian Radioactive Waste Management Program, DOE/RW-0005, DOE Office of Civilian Radioactive Waste Management, June 1985.

DOE, 1986a; NNWSI Project Regulatory Document Manual, NNWSI/88-11, DOE Nevada Operations Office, August 1986 (rev. 6, 1 February 1988).

DOE, 1986b; Environmental Assessment, Yucca Mountain Site, Nevada Research and Development Area, Nevada DOE/RW-0073, DOE Office of Civilian Radioactive Waste Management, May 1986.

DOE, 1986c; Issues Hierarchy for a Mined Geologic Disposal System, DOE/RW-0101, DOE OCRWM Office of Geologic Repositories, September 1986.

DOE, 1986d; Quality Assurance Plan for High-Level Radioactive Waste Repositories, DOE/RW-0095, DOE OCRWM Office of Geologic Repositories, August 1986.

DOE, 1987a; Request for Proposal: Design and Implementation of a Licensing Support System, DOE Office of Civilian Radioactive Waste Management, 11 February 1987.

DOE, 1987b; Issues Hierarchy for a Mined Geologic Disposal System (OGR/B-10), DOE/RW-0101, Revision 1, OCRWM, Office of Geologic Repositories, August 1987.

DOE, 1987c; OCRWM Mission Plan Amendment, DOE/RW-0128, OCRWM, June 1987.

DOE, 1988a; Draft Environmental Regulatory Compliance Plan for Site Characterization of the Yucca Mountain Site, DOE/RW-0177, DOE Nevada Operations Office, January 1988.

DOE, 1988b; Consultation Draft Site Characterization Plan, Yucca Mountain Site, Nevada Research and Development Area, Nevada, DOE/RW-0160, DOE Office of Civilian Radioactive Waste Management, January 1988.

DOE, 1988c; Consultation Draft Site Characterization Plan Overview, Yucca Mountain Site, Nevada Research and Development Area, Nevada, DOE Office of Civilian Radioactive Waste Management, January 1988.

DOE, 1988d; Licensing Support System Preliminary Needs Analysis, DOE/LSS-001, OCRWM, Office of Resource Management, February 1988.

REFERENCES
(continued)

Jordan, 1986; Discovery and Rulemaking Perspectives on the Use of an Information Storage and Retrieval System in the Licensing Proceedings for the High-Level Waste Repository, John S. Jordan & Associates, prepared for NRC, 14 February 1986.

Jordan, 1988; Basic Ideas Underlying Design and Development of an Information System to Support Proceedings before the Nuclear Regulatory Commission on a High-Level Waste Repository, Part 1: General System Concept, John S. Jordan & Associates, prepared for DOE OCRWM, 29 January 1988.

NRC, 1987; 1986 Annual Report, NUREG-1145, Nuclear Regulatory Commission, 15 June 1987.

Olmstead, W.J., 1987; Alternative Dispute Resolution - The NRC's High-Level Waste Negotiated Rulemaking, presented at the ALI-ABA Course of Study, Atomic Energy Licensing and Regulation, Washington, DC, September 14-15, 1987.

SAIC, 1986a; NNWSI Project Information Management System Concepts Evaluation Report, Draft Report, DOE/NV/10270-8, Science Applications International Corp., May 1986.

SAIC, 1986b; Supporting Data and Calculations for the NNWSI Project Information Management Systems Concepts Evaluation Report, DOE/NV/10270-11, Science Applications International Corp. for DOE Nevada Operations Office, December 1986.

Weston, 1985; Licensing Information System Requirements Study: Draft Report of Findings, Roy F. Weston, Inc., prepared for OCRWM Office of Geologic Repositories, 2 October 1985.

Weston, 1986; Generic Requirements for Mined Geologic Disposal System, DOE/RW-0090, Roy F. Weston, Inc., June 1986 (rev. 3, 4 March 1987).

Young, 1987; LSS Functional Requirements and Design Concepts Report, Arthur Young International, prepared for OCRWM, 31 March 1987.

APPENDIX A

DETAILS OF SIZING COUNTS FOR THE SITE CHARACTERIZATION PLAN

Table A. Details of Sizing Counts for the Site Characterization Plan

Section/Topic	Pages	Figures	References	Codes and Regulations
Volume 1				
Organization, Table of Contents, Preface	158	0	0	
Introduction	15	0		
Introduction, References	2		6	3
Part A. Mined Geologic Disposal System	15	0		
Part A. References	3	0	11	1
Chapter 1. Geology	347	96		
Chapter 1. References	72	0	621	3
Chapter 2. Geoengineering	119	26		
Chapter 2. References	19	0	156	1
Volume 2				
Organization, Table of Contents	23	0	0	0
Chapter 3. Hydrology	241	49		
Chapter 3. References	25	0	218	1
Chapter 4. Geochemistry	153	29		
Chapter 4. References	19	0	153	0
Chapter 5. Climatology and Meterology	107	24		
Chapter 5. References	21	0	183	2
Volume 3				
Organization, Table of Contents	26	0		
Chapter 6. Conceptual Design of a Repository	349	94		
Chapter 6. References	23	0	190	8
Chapter 7. Waste Package	241	45		
Chapter 7. References	32		255	6

Table A. Details of Sizing Counts for the Site Characterization Plan (Continued)

Section/Topic	Pages	Figures	References	Codes and Regulations
Volume 4				
Organization, Table of Contents	106	0		
Chapter 8. Introduction (8.0)	11	1		
Chapter 8. Rationale (8.1)	12	1		
Chapter 8. Issues (8.2)	233	15		
Chapter 8. Planned Tests, Analyses, and Studies (8.3)	4	0		
Site Program (8.4)	2	0		
Site Overview (8.3.1.1)	2	0		
Geohydrology (8.3.1.2)	350	38		
Geochemistry (8.3.1.3)	134	16		
Volume 5				
Organization	3	0		
Rock Characteristics	101	17		
Climate (8.3.1.5)	109	8		
Erosion (8.3.1.6)	32	6		
Rock Dissolution (8.3.1.7)	3	0		
Postclosure Tectonics (8.3.1.8)	126	13		
Human Interference (8.3.1.9)	49	7		
Population Density and Distribution (8.3.1.10)	2	0		
Land Ownership and Mineral Rights (8.3.1.11)	4	0		
Meterology (8.3.1.12)	32	4		
Offsite Installation (8.3.1.13)	18	2		
Surface Characteristics (8.3.1.14)	68	9		
Thermal and Mechanical Properties (8.3.1.15)	85	5		

Table A. Details of Sizing Counts for the Site Characterization Plan (Continued)

Section/Topic	Pages	Figures	References	Codes and Regulations
Preclosure Hydrology (8.3.1.16)	27	6		
Preclosure Tectonics (8.3.1.17)	193	16		
Volume 6				
Organization	3	0		
Chapter 8. Repository Program (8.3.2)	2			
Repository Overview (8.3.2.1)	30	2		
Configuration of Underground Facilities Postclosure (8.3.2.2)	93	10		
Repository Design Criteria for Radiological Safety (8.3.2.3)	48	4		
Nonradiological Health and Safety (8.3.2.4)	33	3		
Preclosure Design and Technical Feasibility (8.3.2.5)	116	13		
Seal Program (8.3.3)	2	0		
Seal Overview (8.3.3.1)	7	1		
Seal Characteristics (8.3.3.2)	52	7		
Waste Package Program (8.3.4)	5	1		
Waste Package Overview (8.3.4.1)	8	1		
Waste Package Characteristics, Postclosure (8.3.4.2)	62	6		
Waste Package Production Technologies (8.3.4.4)	7	2		
Performance Assessment Report (8.3.5)	2	0		
Strategy for Preclosure Performance Assessment (8.3.5.1)	21	3		
Waste Retrievability (8.3.5.2)	64	10		

Table A. Details of Sizing Counts for the Site Characterization Plan (Continued)

Section/Topic	Pages	Figures	References	Codes and Regulations
Public Radiological Exposures, Normal Conditions (8.3.5.3)	28	4		
Worker Radiological Safety, Normal Conditions (8.3.5.4)	31	4		
Accidental Radiological Releases (8.3.5.5)	34	6		
Higher Level Findings, Preclosure Radiological Safety (8.3.5.6)	14	1		
Higher Level Findings - Ease and Cost of Construction (8.3.5.7)	17	1		
Strategy for Postclosure Performance Assessment (8.3.5.8)	10	2		
Containment by Waste Package (8.3.5.9)	98	3		
Volume 7				
Organization	2	0		
Engineered Barrier System Release Rates (8.3.5.10)	80	7		
Seal Performance (8.3.5.11)	6	2		
Ground Water Travel Time (8.3.5.12)	70	12		
Total System Performance (8.3.5.13)	106	10		
Individual Protection (8.3.5.14)	15	2		
Ground-Water Protection (8.3.5.15)	13	2		
Performance Confirmation (8.3.5.16)	2	0		
NRC Siting Criteria (8.3.5.17)	99	2		
Higher Level Findings, Postclosure System and Technical Guidelines (8.3.5.18)	29	2		

Table A. Details of Sizing Counts for the Site Characterization Plan (Continued)

Section/Topic	Pages	Figures	References	Codes and Regulations
Completed Analytical Techniques (8.3.5.19)	11	0		
Analytical Techniques Requiring Development (8.3.5.20)	10	0		
Planned Site Preparation Activities (8.4)	76	30		
Milestones, Decision Points and Schedule (8.5)	93	36		
Quality Assurance Program (8.6)	47	3		
Decontamination and Decommissioning (8.7)	5	0		
Chapter 8. References	56	0	499	13
Glossary and Acronyms	120	0		
TOTALS	5355	724	2292	38

APPENDIX B

SUBJECT MATTER CATEGORIES OF DOCUMENTS

FOR INCLUSION OR EXCLUSION IN LSS

The following listing, revised March 25, 1988, describes the categories of documents which the State of Nevada wants to be included in the LSS in searchable full text.

A. The following is a list of subject matter categories of documents, whether "prospective" or "backlog", which should be included in the LSS in searchable full text. The term "document" as used herein, means any written, printed, recorded or graphic matter, however, produced or reproduced, prepared during, or referring or relating to the time period involved in the subject proceeding or concerning or related in whole or in part to any issue or subject matter raised or referred to in the subject proceeding. If a document has been prepared in several copies or additional copies have been made and the copies are not identical, whether by reason of subsequent modification of a copy or by the addition of notations or other modifications, the non-identical copy is a separate document. "Document" as used herein specifically includes writings, statements, depositions, diaries, datebooks, calendars, notes, memoranda, correspondence, files, transcripts of meetings and electronic recordings maintained by any person or group of persons as part of his or their personal or official files, whether at home or at work. "Document" as used herein, includes documents as described above, notwithstanding any claim of privilege with respect to disclosure of such document.

1. Any document pertaining to the location of valuable natural resources, hydrology, geophysics, seismic activity, atomic energy defense activities, proximity to water supplies, proximity to populations, the effect upon the rights of users of water, proximity to components of the National Park System, the National Wildlife Refuge System, the National Wildlife and Scenic River System, the National Wilderness Preservation System, or National Forest Lands, proximity to sites where high-level radioactive waste and spent nuclear fuel is generated or temporarily stored, spent fuel and nuclear waste transportation, safety factors involved in moving spent fuel or high-level nuclear waste to a repository, the cost and impact of transporting spent fuel and nuclear waste to a repository site, the advantages of regional distribution in siting of repositories, and various geologic media in which sites for repositories may be located/

2. Any document related to repository siting, construction, or operation, or the transportation of spent nuclear fuel and high-level nuclear waste, not categorized as an "excluded document", produced by or in the possession of the Los Alamos National Laboratory, the Lawrence Livermore National Laboratory, Sandia National Laboratory, Lawrence Berkeley Laboratory, Oak Ridge National Laboratory, the United States Geologic Survey or any other contractor of the Office of Civilian Radioactive Waste Management in general.

3. All documents related to the physical attributes of the Basin and Range Province of the continental United States.

4. Any document listing and/or considering any site or location other than Yucca Mountain as a possible location for a high-level nuclear waste repository, or any alternative technology to deep geologic disposal.

5. Any document analyzing the effect of the development of a repository at Yucca Mountain on the rights of users of water in the Armagosa ground-water basin in Nevada.

6. Any document analyzing the health and safety implications to the people and environment of the transportation of spent fuel between locations where spent fuel is generated and Yucca Mountain, Nevada, or any other site nominated for repository characterization on May 28, 1986, including, but not limited to:

a. Any analysis of possible human error in the manufacture of spent fuel casks;

b. Any analysis of the actual population density along all of any specific projected routes of travel;

c. Any analysis of releases from any actual radioactive material transportation incidents;

d. Any analysis of the emergency response time in any actual radioactive materials transportation incident;

e. Any actual accident data on any specific projected routes of travel.

f. Any calculations or projections of the probability of accidents on any specific projected routes of travel;

g. Any data on the physical properties or containment capabilities of spent fuel casks which have been used or which are projected to be used at any hypothetical or actual projected repository.

h. Any analysis of modeling of the containment capabilities of spent fuel casks under a stress scenario;

i. Any analysis or comparison of spent fuel casks projected to be used against the spent fuel cask certification standards of the Nuclear Regulatory Commission;

j. Any analysis of the containment capabilities of spent fuel casks containing spent fuel which has been burned up over a extended period.

7. Any document analyzing or comparing Yucca Mountain, Nevada with any other site in the same "geohydrologic setting".

8. Any document relating to any past, present or potential future interference or incompatibility between a Yucca Mountain, Nevada, high-level nuclear waste repository and atomic energy defense activities at the Nevada Test Site.

9. Any document related to the land status, use or ownership or Yucca Mountain, Nevada.

10. Any document considering or analyzing the attributes or detriments or any engineered barrier upon the radioisotope isolation capability of Yucca Mountain, Nevada, or any other site considered.

11. Any document evaluating the effect of extended fuel burn-up on Yucca Mountain, Nevada's adequacy as a repository site for disposal of spent fuel or upon the design of any such theoretical repository.

12. Any document analyzing or investigating the potential for discharge of radioisotopes into the Death Valley National Monument.

13. Any document analyzing the recharge of the underlying saturated zone or the hydroconductivity of the unsaturated zone at Yucca Mountain.

14. Any document containing any data or analysis of volcanic action in the volcanic system of which Yucca Mountain is a part.

15. Any document containing any data or analysis of events of tectonic faulting at Yucca Mountain, either at or beneath the surface of the ground, in tuffacious rock generally, or in the volcanic system of which Yucca Mountain is a part.

16. Any document containing instructions or other limitations on the scope of work to be performed by Department of Energy personnel or contractors' personnel.

17. Any document pertaining to prevention or control of human intrusion at the Yucca Mountain site.

B. The following is a list of documents which may be excluded from the licensing support system notwithstanding their dates of production.

1. Identical copies of documents which are otherwise includable within the searchable full text system.

2. Letters of transmittal used to accompany the transmission of programmatic documents within the Department of Energy. A programmatic document is one related to administration or execution of the Department of Energy's nuclear waste program which contains no reference to any original data, scientific inquiry, site or facility engineering or other data analyses.

3. Documents submitted for reimbursement of personal expenses of travel of Department of Energy personnel or contractor personnel other than for travel to Yucca Mountain.

4. Any documents pertaining exclusively to the management or administration of the U.S. Department of Energy, or the Office of Civilian Radioactive Waste Management. An administrative document is one pertaining to financial management, procurement, personnel, office space, contracting, etc., which does not contain, or refer to, original data, scientific inquiry, transportation data or analysis, engineering data, design or analysis, site analysis or comparison, radioactive or other releases to the environment, cask design or analysis, waste acceptance rate, or the operation of a geologic repository or monitored retrievable storage facility.

APPENDIX C

LICENSING SUPPORT SYSTEM TOPICS

FOR INCLUSION

The following is the Environmental Defense Fund's list of topics for automatic inclusion into the LSS. The transmittal letter with the list is dated March 9, 1988. On March 22, 1988 the listing was revised. The revisions have been incorporated in the following pages.

LICENSING SUPPORT SYSTEM TOPICS FOR INCLUSION

I. The Site

A. LOCATION, GENERAL APPEARANCE AND TERRAIN, AND PRESENT USE

B. GEOLOGIC CONDITIONS
  1. Stratigraphy and volcanic history of the Yucca Mountain area
     a. Caldera evolution and genesis of ash flows
     b. Timber Mountain Tuff
     c. Paintbrush Tuff
     d. Tuffaceous beds of Calico Hills
     e. Crater Flat Tuff
     f. Older tuffs
  2. Structure
  3. Seismicity
  4. Energy and mineral resources
     a. Energy resources
     b. Metals
     c. Nonmetals

C. HYDROLOGIC CONDITIONS
  1. Surface water
  2. Ground water
     a. Ground water movement
     b. Ground water quality
  3. Present and projected water use in the area

D. ENVIRONMENTAL SETTING
  1. Land use
     a. Federal use
     b. Agricultural
        i. Grazing land
        ii. Cropland
     c. Mining
     d. Recreation
     e. Private and commercial development
  2. Terrestrial and aquatic ecosystems
     a. Terrestrial vegetation
        i. Larrea-Ambrosia
        ii. Larrea-Ephedra or Larrea-Lycium
        iii. Colegyne
        iv. Mixed transition
        v. Grassland-burn site
     b. Terrestrial Wildlife
        i. Mammals
        ii. Birds
        iii. Reptiles
     c. Special-interest species
     d. Aquatic ecosystems
  3. Air quality and weather conditions: Air quality
  4. Noise
  5. Aesthetic resources

6. Archaeological, cultural, and historical resources
7. Radiological background
    a. Monitoring program
    b. Dose assessment

E. TRANSPORTATION
1. Highway infrastructure and current use
2. Railroad infrastructure and current use

F. SOCIOECONOMIC CONDITIONS
1. Economic conditions
    a. Nye County
    b. Clark County
    c. Methodology
2. Population density and distribution
    a. Populations of the State of Nevada
    b. Population of Nye County
    c. Population of Clark County
3. Community services
    a. Housing
    b. Education
    c. Water supply
    d. Waste-water treatment
    e. Solid waste
    f. Energy utilities
    g. Public safety services
    h. Medical and social services
    i. Library facilities
    j. Parks and recreation
4. Social conditions
    a. Existing social organization and social structure
        i. Rural social organization and structure
        ii. Social organization and structure in urban Clark County
    b. Culture and lifestyle
        i. Rural culture
        ii. Urban culture
    c. Community attributes
    d. Attitudes and perceptions toward the repository
5. Fiscal and governmental structure

II. EXPECTED EFFECTS OF SITE CHARACTERIZATION ACTIVITIES

A. SITE CHARACTERIZATION ACTIVITIES
1. Field studies
    a. Exploratory drilling
    b. Geophysical surveys
    c. Geologic mapping
    d. Standard operating practices for reclamation of areas disturbed by field studies
2. Exploratory shaft facility
    a. Surface facilities
    b. Exploratory shaft and underground workings
    c. Secondary egress shaft

d. Exploratory shaft testing program
e. Final disposition
f. Standard operating practices that would minimize potential environmental damage

3. Other Studies
   a. Geodetic surveys
   b. Horizontal core drilling
   c. Studies of past hydrologic conditions
   d. Studies of tectonics, seismicity, and volcanism
   e. Studies of seismicity induced by weapons testing
   f. Field experiments in G-Tunnel facilities
   g. Laboratory studies

B. EXPECTED EFFECTS OF SITE CHARACTERIZATION
   1. Expected effects on the environment
      a. Geology, hydrology, land use and surface soils
         i. Geology
         ii. Hydrology
         iii. Land use
         iv. Surface soils
      b. Ecosystems
      c. Air quality
      d. Noise
      e. Aesthetics
      f. Archaeological, cultural, and historical resources
   2. Socioeconomic and transportation conditions
      a. Economic conditions
         i. Employment
         ii. Materials
      b. Population density and distribution
      c. Community services
      d. Social conditions
      e. Fiscal and governmental structure
      f. Transportation
   3. Worker safety
   4. Irreversible and irretrievable commitment of resources

C. ALTERNATIVE SITE CHARACTERIZATION ACTIVITIES

III. REGIONAL AND LOCAL EFFECTS OF LOCATING A REPOSITORY AT THE SITE

A. THE REPOSITORY
   1. Construction
      a. The surface facilities
      b. Access to the subsurface
      c. The subsurface facilities
      d. Other construction
         i. Access route
         ii. Railroad
         iii. Mined rock handling and storage facilities
         iv. Shafts and other facilities
   2. Operations
      a. Emplacement phase
         i. Waste receipt

ii. Waste emplacement
b. Caretaker phase
3. Retrievability
4. Decommissioning and closure
5. Schedule and labor force
6. Material and resource requirements

B. EXPECTED EFFECTS ON THE PHYSICAL ENVIRONMENT
1. Geologic impacts
2. Hydrologic impacts
3. Land use
4. Ecosystems
5. Air quality
a. Ambient air-quality regulations
b. Construction
c. Operations
d. Decommissioning and closure
6. Noise
a. Construction
b. Operations
c. Decommissioning and closure
7. Aesthetic resources
8. Archaeological, cultural, and historical resources
9. Radiological effects
a. Construction
b. Operation
i. Worker exposure during normal operation
ii. Public exposure during normal operation
iii. Accidental exposure during operation

C. EXPECTED EFFECTS OF TRANSPORTATION ACTIVITIES
1. Transportation of people and materials
a. Highway impacts
i. Construction
ii. Operations
iii. Decommissioning
b. Railroad impacts
2. Transportation of nuclear wastes
a. Shipment and routing nuclear waste shipments
i. National shipment and routing
ii. Regional shipment and routing
b. Radiological impacts
i. National impacts
ii. Regional impacts
iii. Maximally exposed individual impacts
c. Nonradiological impacts
i. National impacts
ii. Regional impacts
d. Risk Summary
i. National risk summary
ii. Regional risk summary
e. Costs of nuclear waste transportation
f. Emergency response

D. EXPECTED EFFECTS ON SOCIOECONOMIC CONDITIONS
   1. Economic conditions
      a. Labor
      b. Materials and resources
      c. Cost
      d. Income
      e. Land use
      f. Tourism
   2. Population density and distribution
   3. Community services
      a. Housing
      b. Education
      c. Water supply
      d. Waste-water treatment
      e. Public safety services
      f. Medical services
      g. Transportation
   4. Social conditions
      a. Social structure and social organization
         i. Standard effects on social structure and social organization
         ii. Special effects on social structure and social organization
      b. Culture and lifestyle
      c. Attitudes and perceptions
   5. Fiscal conditions and government structure

IV. SUITABILITY OF THE YUCCA MOUNTAIN SITE FOR DEVELOPMENT AS A REPOSITORY

A. SUITABILITY OF THE YUCCA MOUNTAIN SITE FOR DEVELOPMENT AS A REPOSITORY: EVALUATION AGAINST THE GUIDELINES
1. Technical guidelines
   a. Postclosure site ownership and control
      i. Data relevant to the evaluation
      ii. Favorable condition
      iii. Potentially adverse condition
      iv. Evaluation and conclusion for the qualifying condition on the postclosure site ownership and control guidelines
   b. Population density and distribution
      i. Data relevant to the evaluation
      ii. Favorable conditions
      iii. Potentially adverse conditions
      iv. Disqualifying condition
      v. Evaluation and conclusion for the qualifying condition on the population density and distribution guideline
   c. Preclosure site ownership and control
      i. Data relevant to the evaluation
      ii. Favorable condition
      iii. Potentially adverse condition
      iv. Evaluation and conclusion for the qualifying condition on the preclosure site ownership and control guideline
   d. Meteorology
      i. Data relevant to the evaluation
      ii. Favorable condition

iii. Potentially adverse condition
iv. Evaluation and conclusion for the qualifying condition on the meterology guideline

e. Offsite installations and operations
   i. Data relevant to the evaluation
   ii. Favorable conditions
   iii. Potentially adverse conditions
   iv. Disqualifying condition
   v. Evaluation and conclusion for the qualifying condition on the offsite installations operations guideline

f. Environmental quality
   i. Data relevant to the evaluation
   ii. Favorable conditions
   iii. Potentially adverse conditions
   iv. Disqualifying conditions
   v. Evaluation and conclusion for the qualifying condition on the environmental quality guidelines

g. Socioeconomic impacts
   i. Data relevant to the evaluation
   ii. Favorable conditions
   iii. Potentially adverse conditions
   iv. Disqualifying condition
   v. Evaluation and conclusion for the qualifying condition on the socioeconomic guideline

h. Transportation
   i. Data relevant to the evaluation
   ii. Favorable conditions
   iii. Potentially adverse conditions
   iv. Disqualifying condition
   v. Evaluation and conclusion for the qualifying condition on the transportation guideline

2. Preclosure System
   a. Preclosure system: radiological safety
      i. Data relevant to the evaluation
      ii. Evaluation of the Yucca Mountain site
      iii. Conclusion for the qualifying condition on the preclosure system guideline radiological safety
   b. Preclosure system: environment, socioeconomics, and transportation

3. Postclosure technical
   a. Geohydrology
      i. Data relevant to the evaluation
      ii. Favorable conditions
      iii. Potentially adverse conditions
      iv. Disqualifying condition
      v. Evaluation and conclusion for the qualifying condition on the postclosure geohydrology guideline
   b. Geochemistry
      i. Data relevant to the evaluation
      ii. Favorable conditions
      iii. Potentially adverse conditions
      iv. Evaluation and conclusion for the qualifying condition on the postclosure geochemistry guideline
      v. Plans for site characterization

- c. Rock characteristics
    - i. Data relevant to the evaluation
    - ii. Favorable conditions
    - iii. Potentially adverse conditions
    - iv. Evaluation and conclusion for the qualifying condition on the postclosure rock characteristics guideline
- d. Climate changes
    - i. Data relevant to the evaluation
    - ii. Favorable conditions
    - iii. Potentially adverse conditions
    - iv. Evaluation and conclusion for the climate changes qualifying condition
- e. Erosion
    - i. Data relevant to the evaluation
    - ii. Favorable conditions
    - iii. Potentially adverse conditions
    - iv. Disqualifying condition
    - v. Qualifying condition
- f. Dissolution
    - i. Data relevant to the evaluation
    - ii. Favorable conditions
    - iii. Potentially adverse conditions
    - iv. Disqualifying condition
    - v. Evaluation and conclusion for the qualifying condition on the postclosure and dissolution guideline
- g. Tectonics
    - i. Data relevant to the evaluation
    - ii. Favorable conditions
    - iii. Potentially adverse conditions
    - iv. Disqualifying condition
    - v. Evaluation and conclusion for the qualifying condition on the postclosure tectonics guideline
- h. Human interference: natural resources and site ownership and control
    - i. Data relevant to the evaluation
    - ii. Favorable conditions
    - iii. Potentially adverse conditions
    - iv. Disqualifying condition
    - v. Evaluation and conclusion for the qualifying condition on the postclosure human interference and natural resources technical guideline

4. Postclosure system
    - a. Evaluation of the Yucca Mountain site
        - i. Quantitative analyses
        - ii. Qualitative analysis
    - b. Summary and conclusion for the qualifying condition on the postclosure system guideline
5. Preclosure technical
    - a. Surface characteristics
        - i. Data relevant to the evaluation
        - ii. Favorable conditions
        - iii. Potentially adverse conditions
        - iv. Evaluation and conclusion for the qualifying condition on the preclosure surface characteristics guideline

b. Rock characteristics
  i. Data relevant to the evaluation
  ii. Favorable conditions
  iii. Potentially adverse conditions
  iv. Disqualifying condition
  v. Evaluation and conclusion for the qualifying condition on the preclosure rock characteristics guideline

c. Hydrology
  i. Data relevant to the evaluation
  ii. Favorable conditions
  iii. Potentially adverse conditions
  iv. Disqualifying condition
  v. Evaluation and conclusion for the qualifying condition on the preclosure hydrology guideline

d. Tectonics
  i. Data relevant to the evaluation
  ii. Favorable conditions
  iii. Potentially adverse conditions
  iv. Disqualifying condition
  v. Evaluation and conclusion for the qualifying condition on the preclosure tectonics guideline

6. Ease and cost of siting, construction, operation, and closure
  a. Data relevant to the evaluation
  b. Evaluation
  c. Conclusions for the qualifying condition on the ease and cost of siting, construction, operation, and closure guideline

7. Conclusion regarding suitability of the Yucca Mountain Site for a repository

B. PERFORMANCE ANALYSES

1. Preclosure radiological safety assessments
  a. Preclosure radiation protection standards
  b. Methods for preclosure radiological assessment
    i. Radiological assessment of construction activities
    ii. Radiological assessment of normal operations
    iii. Radiological assessment of accidental releases

2. Preliminary analysis of postclosure performance
  a. Subsystem descriptions
    i. Engineered barrier subsystem
    ii. The natural barrier subsystem
  b. Preliminary performance analyses of the major components of the system
    i. The waste package lifetime
    ii. Release rate from the engineered barrier subsystem
  c. Preliminary system performance description and analysis
  d. Comparisons with regulatory performance objectives
  e. Preliminary evaluation of disruptive events: disruptive natural processes
  f. Conclusions

V. TRANSPORTATION

A. REGULATIONS RELATED TO SAFEGUARDS

1. Safeguards
2. Conclusion

B. PACKAGINGS
   1. Packaging design, testing, and analysis
   2. Types of packaging
      a. Spent fuel
      b. Casks for defense high-level waste and West Valley high-level waste
      c. Casks for use from an MRS to the repository
   3. Possible future developments
      a. Mode-specific regulations
      b. Overweight truck casks
      c. Rod consolidation
      d. Advanced handling concepts
      e. Combination storage/shipping casks

C. POTENTIAL HAZARDS OF TRANSPORTATION
   1. Potential consequences to an individual exposed to a maximum extent
      a. Normal transport
      b. Accidents
   2. Potential consequences to a large population from very severe transportation accidents
   3. Risk Assessment
      a. Outline of method for estimating population risks
      b. computational models and methods for population risks
      c. Changes to the analytical models and methods for population risks
      d. Transportation scenarios evaluated for risk analysis
      e. Assumption about wastes
      f. Operational considerations for use in risk analysis
      g. Values for factors needed to calculate population risks
      h. Results of population risk analyses
      i. Uncertainties
   4. Risks associated with defective cask construction, lack of quality assurance, inadequate maintenance and human error

D. COST ANALYSIS
   1. Outline method
   2. Assumptions
   3. Models
   4. Cost estimates
   5. Limitations of results

E. BARGE TRANSPORT TO REPOSITORIES

F. EFFECT OF A MONITORED RETRIEVABLE STORAGE FACILITY ON TRANSPORTATION ESTIMATES

G. EFFECT OF AT-REACTOR ROD CONSOLIDATION ON TRANSPORTATION ESTIMATES

H. CRITERIA FOR APPLYING TRANSPORTATION GUIDELINE

I. DOE RESPONSIBILITIES FOR TRANSPORTATION SAFETY
   1. Prenotification
   2. Emergency response
   3. Insurance coverage for transportation accidents

J. MODEL MIX
   1. Train shipments
      a. Ordinary
      b. Dedicated train
   2. Truck shipments
      a. legal weight
      b. overweight

K. INFRASTRUCTURE AND CURRENT USE
   1. Rail
   2. Highway

APPENDIX D

U.S. DEPARTMENT OF ENERGY'S LISTING OF MATERIAL TO BE INCLUDED OR EXCLUDED FROM THE LSS

The following listings, prepared by DOE staff, are as revised on March 25, 1988.

LICENSING SUPPORT SYSTEM

ENHANCED FULL TEXT

INCLUDABLE LIST FOR ALL PARTIES

PROPOSED BY THE DEPARTMENT OF ENERGY

Relevant records of all parties (except those where an appropriate privilege applies) related to the HLW licensing decision to be placed in the LSS in enhanced full text.

- Technical reports and analyses including those developed by contractors
- QA/QC records including qualification and training records
- External correspondence
- Internal memoranda
- Meeting minutes, including DOE/NRC meetings, Commission meetings
- Drafts (i.e., those submitted for decision beyond the first level of management or similar criterion)
- Congressional Q's & A's
- "Regulatory" documents related to HLW site selection and licensing, such as:
    - Draft and final environmental assessments
    - Site Characterization Plans
    - Site Characterization progress reports
    - Issue resolution reports
    - Rulemakings
    - Public and agency comments on documents
    - Response to public comments
    - Environmental Impact Statement, Comment Response Document, and related references
    - License Application (LA), LA data base, and related references
    - Topical reports, data, and data analysis
    - Recommendation Report to President
    - Notice of Disapproval, if submitted

LICENSING SUPPORT SYSTEM

EXCLUDABLE LIST FOR ALL PARTIES

PROPOSED BY THE DEPARTMENT OF ENERGY

MATERIAL NOT TO BE INCLUDED IN THE LICENSING SUPPORT SYSTEM

Public Domain Material

- Official notice material, such as encyclopedias, dictionaries, and text books
- Bulk/public correspondence
- General press clippings, periodicals, press releases, and circulation/direct distribution mail
- Procurement material, such as purchase orders, contract records, and RFP's other than scope of work (some material may also be privileged)
- Staff speeches & publications
- Report references

Personal Records

- Personal mail and other personal material

MATERIAL INCLUDED IN THE LSS BUT NOT IN ENHANCED FULL TEXT

Unsuitable Form

- Raw data
- Computer runs
- Computer programs & codes
- Field notes
- Maps & photographs
- Core samples

Privileged Material

- Information protected by FOIA and Privacy Act
- Classified material including Safeguards and Security Information
- Personnel records, travel requests and vouchers

APPENDIX E

ABBREVIATIONS USED

ABBREVIATIONS USED

ACD	Advanced Conceptual Design
ACRS	NRC Advisory Committee on Reactor Safeguards
ANSI	American National Standards Institute
ARS	DOE Automated Records System
ASLB	NRC Atomic Safety and Licensing Board
ASME	American Society of Mechanical Engineers
CDSCP	Consultant Draft SCP
CFR	Code of Federal Regulations
CRD	Conceptual Repository Design
DEIS	Draft Environmental Impact Statement
DOE	Department of Energy
EPA	Environmental Protection Agency
FEIS	Final Environmental Impact Statement
FOIA	Freedom of Information Act
HLW	High Level Waste
LA	License Application
LAD	License Application Design
LSS	Licensing Support System
MRS	Monitored Retrievable Storage
NNWSI	Nevada Nuclear Waste Storage Investigations
NRAC	Negotiated Rulemaking Advisory Committee, officially known as the HLW Licensing Support System Advisory Committee
NRC	Nuclear Regulatory Commission
NTS	Nevada Test Site
NWPA	Nuclear Waste Policy Act of 1982
OCRWM	DOE Office of Civilian Radioactive Waste Management

ABBREVIATIONS USED
(continued)

OGR	OCRWM Office of Geologic Repositories
OMB	Office of Management and Budget
PSAR	Preliminary Safety Analysis Report
QA	Quality Assurance
QC	Quality Control
RFP	Request For Proposal
SAIC	Science Applications International Corporation
SCA	Site Characterization Analysis
SCP	Site Characterization Plan
SCPCD	Site Characterization Plan Conceptual Repository Design
SER	Safety Evaluation Report
UIC	Underground Injection Control
USC	The United States Code
WP	Waste Package

LICENSING SUPPORT SYSTEM
CONCEPTUAL DESIGN ANALYSIS

SCIENCE APPLICATIONS INTERNATIONAL CORPORATION

TABLE OF CONTENTS

TABLE OF CONTENTS
(continued)

Page

TABLE OF CONTENTS
(continued)

Page

TABLE OF CONTENTS
(continued)

Preface

This is the third in a series of four reports on the Licensing Support System (LSS) prepared by the DOE Office of Civilian Radioactive Waste Management (OCRWM) for the Office of Management and Budget (OMB). The LSS is an information management system intended to support the needs of all the parties involved in repository licensing, including the Department of Energy (DOE) and the Nuclear Regulatory Commission (NRC). These reports are:

Preliminary Needs Analysis

Preliminary Data Scope Analysis

Conceptual Design Analysis

Benefit-Cost Analysis

The Preliminary Needs Analysis, issued in February 1988, and the Preliminary Data Scope Analysis, issued in March 1988, constitute the system requirements basis for developing a Conceptual Design, presented in this report. The Benefit-Cost Analysis evaluates alternatives within this conceptual design. These four reports, and subsequent refinements, are intended to provide the basis for determining the LSS design specifications.

Note that Appendix B contains a revised version of Table 8 (Projected size of the LSS data base, 1990 - 2009) of the Preliminary Data Scope Analysis. Revisions have been made for the estimates of pages added during 1993 and 1998, based on a re-evaluation of the expected levels of activity consistent with the methods described in that report.

Executive Summary

This report analyzes the design requirements for the Licensing Support System (LSS) and presents viable design concepts and implementation approaches that satisfy the expected functional requirements of LSS users. It presents a detailed conceptual design base that is highly responsive to these requirements and has a low associated development risk.

The Preliminary Needs Analysis and Preliminary Data Scope Analysis have defined the requirements of an automated computer-based information storage and retrieval system that must accommodate millions of documents. Based on these studies and the directions perceived from the LSS negotiated rulemaking process, a conceptual design has been formulated which meets the rulemaking requirements. The conceptual design has the following major features:

1) Headers and searchable full text of all documents suitable for inclusion in LSS

2) Bit-map images of all documents in LSS
   - reproduction of documents for quick distribution from central location
   - on-line display and local printing at special workstations

3) Centralized search system and on-line optical disk image system in Washington, DC or Las Vegas, NV

4) Multiple capture systems for:
   - scanning
   - text conversion
   - correction
   - cataloging

5) Workstations capable of displaying readers, ASCII text and images

6) Support for workstations displaying headers and ASCII text only

7) Retrieval through structured index searching of cataloging information and software full text searching of documents

8) Electronic mail

In analyzing the LSS requirements and formulating the conceptual design presented, a number of alternatives were examined. Some were rejected due to a low probability of success. Others not only offered a reasonable risk but also were potentially more cost effective, although they may not meet all of the "non-firm" requirements identified in the Preliminary Needs Analysis. These alternatives were identified as variants to the base and will be subject to further technical and economic investigation.

The variants examined differed from the base conceptual design in the following ways:

I. Two search and image systems replicating the data base, rather than one, located in Washington, DC and Las Vegas

II. Hardware full text search, rather than software

III. No workstations capable of displaying images

IV. Microform digitization rather than optical disk storage of images

V. Off-line microform printing rather than on-line bit-mapped image system

VI. Re-keying text rather than text conversion from scanned bit-map.

VII. Combination of 3, 5 and 6 above.

The base conceptual design and variants are consistent with the requirements identified to date, including the deliberations of the Negotiated Rulemaking Advisory Committee. Indeed, the rulemaking activities have not yet imposed any requirements on the design which were not anticipated in the Preliminary Needs Analysis. It is further not expected that the rulemaking activities will result in any requirements which cannot be met by the base design or one of the variants; however, the possibility still exists that design refinements may be required to reflect changing requirements. It is important to note that the need for the Tracking Subsystems has not been identified to date by the Negotiated Rulemaking Advisory Committee, but rather arises from the Preliminary Needs Analysis discussions with potential DOE (and contractor) users and from the LSS Design and Implementation Contract (DOE, 1987) requirements. For this reason, it is probable that the Tracking Subsystems described here shall be removed from the LSS design and implemented for DOE only.

Significant to the refinement of the design is the feedback from potential users and the benefit-cost analysis. All of the dialogue conducted to this point with potential users of the LSS has been in the absence of a common reference design. This has caused much difficulty in communication, because while one must speak from a certain frame of reference, it is almost assuredly not the same as that of the other party. With this study, a reference point is established for the LSS conceptual design that will facilitate future deliberations. Refinement of the design can benefit from additional experiences of and feedback from potential users, especially in the area of features available to the user. With this base frame of reference, the potential user can now envision certain scenarios of LSS applications and can fill in the various details necessary to more accurately predict system sizing and response.

The Benefit-Cost Analysis, the next report in the series, will permit the investigation of the variants from the standpoint of cost (both savings and additions), and the associated benefits (or lack thereof) of each variant. Such information will provide an economic basis for decisions on how best to meet the various requirements on the system.

The opportunity is now available for substantive input to the design process, which is both encouraged and necessary to refine the process further.

1.0 INTRODUCTION

This report analyzes the design requirements for the Licensing Support System (LSS) and presents viable design concepts and implementation approaches that satisfy the expected functional requirements of LSS users. It presents a detailed conceptual design base that is highly responsive to these requirements and has a low associated development risk.

1.1 Purpose and Scope

The first two reports in this series (Preliminary Needs Analysis and Preliminary Data Scope Analysis) have been a first effort under the LSS Design and Implementation Contract toward developing a sound requirements foundation for the LSS design. It is clear that there is not one unique design which can meet these requirements and the detailed specifications to be refined from them. Rather, a variety of basic concepts and myriad variations on those concepts can be acceptable.

The purpose of this analysis is to develop one such concept, for which a benefit-cost analysis will be performed in the last report in this series. Although a number of general LSS concepts have been proposed, the conceptual design presented here is the most detailed to date. In order to present this level of detail, the basic concepts examined needed to be limited. The basic concept presented here (the base conceptual design presented in Section 4.1) was selected because it is highly responsive to the requirements identified and has a low associated development risk.

Section 2.0 specifies the major functions required of LSS and discusses how these functions can be implemented. The basic concept of LSS operation is presented in Section 3.0 and covers both information capture and information retrieval. Section 4.1 describes the proposed base conceptual design in detail, and Section 4.2 describes seven major variants on this design. These variants have been selected to examine significant technological or operational options in implementing LSS functions. These variants will be examined along with the Base Conceptual Design in the forthcoming Benefit-Cost Analysis.

1.2 LSS Subsystems

The Licensing Support System (as described in the LSS Design and Implementation contract RFP) consists of four subsystems, grouped into two categories - Text Storage and Tracking. Each component provides on-line access to a particular type of information.

1.2.1 Text Storage

The Text Storage category has two subsystems: the Records Access Subsystem and the Regulation Access Subsystem. Each subsystem is a library of documents generated by or related to primarily the OCRWM mined geologic repository program and its license application process. The full text of the documents in each will be stored, indexed and made available for on-line search and retrieval. A set of catalog data containing information about each document (such as author(s), title, date, document type, and accession number) will be maintained, indexed and made available for on-line access.

1.2.1.1 Records Access Subsystem

The Records Access Subsystem will primarily contain documents generated by OCRWM program participates. Users of the subsystem will include attorneys, licensing staff, project engineers, and designers. Hard copies or microform copies of all documents in the subsystem will be kept in the LSS Document Archives. The Records Access Subsystem will also contain cataloging data for non-text programmatic records (such as material samples, field instrument strip charts and design drawings). The non-text records will be kept in the LSS Satellite Archives. The Records Access Subsystem cataloging data for each of these items will contain its location in the LSS Archive storage facilities.

1.2.1.2 Regulations Access Subsystem

The Regulations Access Subsystem will contain documents that impose legal requirements on the program, such as, the Nuclear Waste Policy Act (NWPA) of 1982 and its Amendments Act, 10 CFR Part 60, other Federal and state regulations, regulatory guides, and DOE agreements with regulatory agencies. Both structured cataloging data and full text will be stored and available for search and retrieval. Each of the user groups identified in the Preliminary Needs Analysis are expected to use this subsystem.

1.2.2 Tracking

The LSS will also include two Tracking subsystems: the Issues Tracking Subsystem and the Commitments Tracking Subsystem. These subsystems will be used primarily by OCRWM and NRC licensing staff, project management personnel and technical staff to record and review identified and agreed to issues and commitments and the status of progress towards their compliance or completion. In both subsystems the data will be stored in a structured form, similar to the catalog data in the Records Access and Regulations Access Subsystems.

It is important to note that the need for the Tracking Subsystems has not been identified to date by the Negotiated Rulemaking Advisory Committee, but rather arises from the Preliminary Needs Analysis discussions with potential DOE (and contractor) users and from the LSS Design and Implementation Contract (DOE, 1987) requirements. For this reason, it is probable that the Tracking Subsystems described here shall be removed from the LSS design and implemented for DOE only.

1.2.2.1 Issues Tracking Subsystem

Issues are identified efforts that clarify what part of a regulation applies to the program and how compliance is to be achieved. Issues can be raised by DOE or a regulatory agency and resolution is accomplished through research, investigation and negotiation, the results of which are documented and stored in the Records Access Subsystem. The Issues Tracking Subsystem will provide the capability to record and track the resolution of issues in a controlled, automated fashion. Some uses of the subsystem will be by:

1) OCRWM licensing staff who need to track licensing and regulatory related issues for the geologic repository program

2) NRC and other non-DOE parties identified in the Nuclear Waste Policy Act, who are sources of issues

3) OCRWM Site Characterization staff and NNWSI project staff who will follow progress towards resolving the issues in the NNWSI Site Characterization Plan.

1.2.2.2 Commitments Tracking Subsystem

The Commitments Tracking Subsystem is similar to the system described above, but records and tracks the progress of:

1) Commitments made by OCRWM to other organizations

2) Commitments made to OCRWM by other organizations

3) Commitments or action items internal to the OCRWM program.

2.0 REQUIRED LSS FUNCTIONS AND THEIR CONCEPTUAL IMPLEMENTATIONS

The LSS will be used to capture, store, access, and present (output) all records, regulations and tracking information that is deemed relevant to obtaining the necessary licenses and permits for the siting, design, construction, operation, and closure of a geologic repository for the disposal of spent nuclear fuel and high-level nuclear waste as authorized by the Nuclear Waste Policy Act of 1982 and the Nuclear Waste Policy Amendments Act of 1987. This section describes needed LSS functions and features in these areas that were identified in the Preliminary Needs and the Preliminary Data Scope Analyses, and presents possible implementations. Additional communications, electronic mail and management functions necessary to LSS are also described and evaluated here.

2.1 Capture of Information into the LSS

The LSS must be able to accept a variety of inputs from various sources. Information in the Tracking Subsystems and header information in the Records and Regulations Access Subsystems is created specifically for the LSS and must be keyed into the system. According to the Preliminary Data Scope Analysis, the vast majority of the backlog material appropriate for inclusion in the Records and the Regulations Access Subsystems exists in the form of hardcopy and microform documents. The text of hardcopy documents, needed for the full-text search capability which is expected to be required by the Negotiated Rulemaking Advisory Committee (NRAC), may be entered by re-keying or by the conversion of a scanned image of the document into a text file. Microform information must be converted to and entered as hardcopy or must be directly digitized and similarly converted. The NRAC is also expected to require images of all material in the LSS. The system is also expected to support electronic mail among users (Preliminary Needs Analysis) and will probably be required to accept material from the electronic mail system as input to the Records Access Subsystem. The following sections present the capture functions and features required to support data entry into these subsystems.

The Preliminary Needs Analysis also indicates that rigorous quality control (QC) in the capture process is essential to the usefulness of the LSS. Such QC would not be limited to assuring the conformity of captured material with the original, but must also include the elimination of duplicate input and the verification of cataloging information.

2.1.1 Capture of Tracking Information

2.1.1.1 Issues Tracking

The Issues Tracking function requires the following types of input:

1) Input Issue-Related Information
LSS must provide for the input of issues which relate to NRC, DOE, other federal, state, and local rules and regulations.

2) Input Information Related to the Resolution of Issues
LSS must provide for the input of the description of the resolution of issues. Each resolution must be related to the relevant issue record.

3) Input Information Related to the Status of Issues
LSS must provide for the input of milestones and events related to the handling of an issue.

4) Input Issue-Related Work Plans
LSS must provide for the input of the action plans for issue resolution and responsibilities for attaining resolution of each issue.

5) Input References to Information Related to the Issue
LSS must provide for the input of references to the source documents that initiated or identified the issue, regulations that impact the issue, and information related to the resolution of the issue.

The input of Issues Tracking related textual information will be from LSS terminals (local and remote). The data will be entered from terminal keyboards *via* on-line, fill-in-the-blank type of screens. The procedures that will govern the selection and specification (format) of the data to be entered, as defined above, will contain data entry forms for the transcription of the data to be entered. Each Issues record will contain an unique accession number for identification and retrieval purposes. The data entry procedures will also provide for the edit/changes of text in existing Issues records in the data base. Markups of text in existing documents may be attached to the data entry form to minimize the amount of transcription required.

2.1.1.2 Commitment Tracking

The Commitment Tracking function requires the following types of input:

1) Input Information Related to Commitments
LSS must provide for the input of commitments made to federal agencies, regulatory agencies, states, tribes, and legal parties identified in the Nuclear Waste Policy Act.

2) Input the Resolution of Commitments
LSS must provide for the input of the description of the resolution

of the commitment. Each resolution must be related to the relevant commitment record.

3) <u>Input Information on the Status of Commitments</u>
LSS must provide for the input of milestones and events related information on the status of a commitment.

The input of Commitment Tracking related textual information will be from LSS terminals (local and remote). The data will be entered from terminal keyboards *via* on-line, fill-in-the-blank type of screens. The procedures which will govern the selection and specification (format) of the data to be entered, as defined above, will contain data entry forms for the transcription of the data to be entered. Each Commitment record will contain an unique accession number for identification and retrieval purposes. The data entry procedures will also provide for the edit/change of text in existing Commitment records in the data base.

2.1.2 Capture of Records and Regulations Information

Capture of the LSS records and regulations consists of the processes of scanning documents to obtain the electronic (bit-mapped) image, obtaining text in electronic form for the purpose of preparing for full-text search, cataloging the document for retrieval, and the associated quality control. Due to the large amount of data to be processed, capture of the LSS documents and regulations, and their corresponding cataloging data, is expected to be accomplished at document capture centers located in the vicinity of sites which contain significant backlog or generate significant new documents.

2.1.2.1 Image Capture (Scanning)

Scanning of hardcopy pages to capture the electronic image is required for pages which are to be submitted to the character recognition process and for pages which are not in text form such as graphics, maps, and figures. In addition, since it is required to store the image of the document for the purpose of providing hardcopy on request, it may be reasonable to store those images in electronic form.

2.1.2.2 Text Capture

ASCII text of documents may be captured by the system by one of the following three methods:

1) Direct keying of document text from terminal keyboards
2) Optical character recognition (OCR)
3) Input of ASCII word processor files.

2.1.2.2.1 Direct Keying of Document Text

The LSS should provide the capability to capture document text in ASCII format by direct keying of document text from terminal keyboards. This capability is particularly important for documents that cannot be efficiently converted through the Optical Character Recognition (OCR) process, for example, barely legible copies.

2.1.2.2.2 Optical Character Recognition

The OCR process converts an electronic (bit-mapped) image of a page into ASCII text (a standard pattern for each character and punctuation). The quality of the text produced is highly dependent on the quality of the image that is submitted to the process, i.e. an original printed page with uniform type will produce better results than a fourth generation photocopy with smudges and extraneous markings. Current generation OCR devices can produce text with 99.5% to 99.9% accuracy under optimum conditions. Note that this would still result in 3 to 15 errors in a 3000 character page.

Correction of errors is a manual process, although software tools such as spelling checkers can assist. (A nontrivial consideration is whether or not to correct spelling errors in the original text.) The necessity to correct the errors is dependent on their magnitude and other factors such as:

1) The effect of the errors on full-text retrieval
2) The use of the ASCII text in reading or browsing the document
3) The use of the ASCII text for downloading and file transfer.

The advantage of the OCR process is that it is relatively automated and can be performed without much human intervention up to the point of review and correction. Continuous improvements are being made in OCR technology that will increase speed of production and reduce the error rate. Presently OCR of an image made from scanning a good quality paper copy can be reasonably performed, however OCR from an image produced by blow-back of a microfiche or microfilm is not considered feasible.

2.1.2.2.3 Input of Standard Word Processor Files

It is expected that much of the future documentation for the repository design and licensing will be prepared with word processing equipment. Therefore it is important that the LSS should provide the capability to capture document text directly in electronic form. In order to accommodate the various software programs, a standard input format utilizing the ASCII form (no special codes or printer control characters) will be required. The permissible media include floppy disks and magnetic tape. In order that the image may be captured, a good quality hardcopy must also be provided with the ASCII file.

The major problem with receiving data in machine readable format is the quality assurance of such inputs. It is necessary that the machine readable version of the document be verified as a true representation of the hardcopy. (In many cases last minute changes to a document are made on a typewriter.) Therefore, the quality assurance procedures for the input and verification of word processor inputs to the LSS must be clearly defined and enforced to ensure data base integrity.

2.1.2.2.4 Electronic Mail Input

The LSS is required to serve as a mechanism for the electronic transmission of filings by the parties during the high-level waste proceeding by means of an Electronic mail (E-Mail) capability. These filings must also be captured as records at the request of the sender. This can be accomplished by having the sender transmit a copy of the message along with "fill-in-the-blanks" header information to an LSS mailbox at a document capture workstation. The E-mail message can then be processed through the same input, cataloging, and quality assurance procedures as any other document, except that an image of the message will not exist.

2.1.2.3 Cataloging

The LSS should provide for the capture of catalog information, input by terminal keyboards, that will contain bibliographic data and keywords to support the user access to the documents stored in the Records or Regulations Access Subsystem. The catalog information will be stored in a "header" for each record. The header is created by extracting some of the information in the record and placing it in appropriate fields stored at the beginning or "head" of the record (descriptive cataloging). In addition, the header can include information about the subject content of the record (subject cataloging). A header describing a document can also be entered into a system without any additional information (such as the text of a document or an abstract). Examples of kinds of information obtained during cataloging are:

1) Descriptive information (extracted from a document)
   Author's name and affiliation
   Date of the document
   Title of the document
   Publication information (journal name, publisher, etc.)
   Recipient
   Abstract (if part of the document)

2) Subjective information (assigned to a document)
   Subject category
   Degree of technical detail
   Descriptors or keywords
   Abstract (if written by cataloger)
   Pointer to relevant LSS Issues and Commitments entries

Cataloging can also be computer assisted. Examples of computer-assisted cataloging are assigning an accession number to the document, extracting heavily used words for use as keywords, etc. Cataloging results are not necessarily limited to the header, as in the case where markers of special information are embedded in the text of a document file to facilitate retrieval. Since the primary purpose of cataloging is to provide access points for retrieving (finding) the document in the data base without having to examine sequentially each document in the collection, the development of the header format, i.e., determining which fields are to be included in the header, is a function of both the cataloging process and the retrieval requirements.

The primary tool for the cataloging process is the cataloging manual. The cataloging manual consists of rules or conventions for entering data into each field. These rules or conventions govern either the form or the content of each entry. Generally, rules of form standardize the format in which data is entered, e.g., dates are to be entered in the format yy/mm/dd. Each field in the LSS header will have rules of form. These rules will be part of the cataloging manual. Rules governing the content of the entry are usually contained in controlled vocabularies. A controlled vocabulary or authority list is simply a list of those words available for entry into a field. Words or terms not contained in the list are unacceptable. Controlled vocabularies carry standardization into the content of the field entry by specifying the correct word form(s) of names, subjects, etc., to be used in the data base. Fields such as agency name, corporate affiliation, or document type will have a controlled vocabulary as part of the cataloging manual. These lists will remove any doubt whether to use DOE, Dept. of Energy, Department of Energy, Energy Dept., U.S. Department of Energy, etc. for that government agency. Similarly, the decision to use the term letter, memo, correspondence, or mail for a particular document type will be resolved by consulting the controlled vocabulary for that field.

Controlled vocabularies are also used for subject cataloging. The LSS would use a thesaurus as the controlled vocabulary for terms describing the content of each record. A thesaurus is a controlled vocabulary list that shows the relationships between words or terms in the vocabulary. The relationships are shown as broader, narrower, or related terms. In addition, "use" and "used for" terms are also included. As an example, the broader term for "automobile" is "vehicle", a narrower term is "four-door sedan", and a related term is "motorcycle". A "used for" term would be "car", while the entry for "car" would state "use automobile". Thesaurus entries also contain notes or definitions to explain the application of the term, e.g. "hearing - a legal procedure" (as opposed to "hearing - a physiological process"). Thus a thesaurus would eliminate uncertainties on the part of the cataloger and the user as to which terms to use for cataloging and retrieval. The thesaurus would be arranged both alphabetically and hierarchically to assist the cataloger and the user in describing documents or non-documents consistently, logically, and at an appropriate level of detail. Instructions on how to use the thesaurus and how to submit recommended terms would be included in the thesaurus. Procedures for maintaining, updating, and revising all controlled vocabularies will be included in the cataloging manual.

One of the most significant conclusions of the Preliminary Needs Analysis is that appropriate and extremely high quality cataloging is essential to guarantee the usefulness of the Records and Regulations Access Subsystems. The quality of the cataloging directly determines the usefulness of structured index access to the LSS, which, in turn, is a critical (if not the most critical) technique for identifying material in the system. Much of this cataloging must be performed by highly trained and experienced personnel, particularly the subject cataloging. The development of an appropriate header format and the design of rigorous and reliable cataloging procedures are therefore key aspects of the detailed LSS design.

2.1.3 Capture Quality Control

The capture quality control process should include not only verification of correctness and of fidelity of free text and images to the original material, but also explanation of anomalies or relationships that might surprise an LSS user.

Quality control is accomplished by:

1) Use of an integrated production process that minimizes opportunity for introduction of error
2) A production process that focuses human attention and uses human time efficiently
3) Human follow-up and spot checks
4) Use of automated tools to verify data, enforce consistency, and identify anomalies.

Quality control of imaged data includes:

1) Verification that all pages are present and correctly sequenced
2) Verification that imaged pages are legible, complete, and unskewed
3) Entry of a standard indicator for missing pages or oversized pages.

Software tools to support the image quality control process might include:

1) Tools to re-sequence pages
2) Tools to delete pages.

General policies regarding trade-offs of fidelity to source material versus utility of the data must be articulated and agreed to before the quality control process is implemented. One policy might be to ensure that all data stored in headers complies with standards for spelling and abbreviation, but document text is to reflect spelling, including errors, as they appear in the source document.

Consistency of assignment and of data representation is critical to the usability of catalog data. Catalogers will be highly trained and will have specific, written guidelines available for all fields that require a specific format or that require evaluation and judgment. A variety of software tools support entry and quality control of catalog data. They include:

1) Menus accessing authority lists legal values for fields having controlled vocabularies (<u>e.g.</u>, journal title, document type, author affiliation) to provide on-line tools to help in validity checking

2) Range checks, date conversion checks, and format checks for applicable fields

3) Cut-and-paste capabilities for data such as author names and document titles and importing from windows for data such as descriptors and major subjects.

As the data base grows, far more sophisticated computer assistance can be developed. For example, author names can be checked for similarity to previously-stored data. Names that differ by only a few characters may be presented for human verification. Similarly, publication dates might be verified for feasibility with respect to other publications by the same author. Computer-assisted duplicate detection is required in the early phases of population. Later, automated duplicate detection, perhaps with human verification, may be more practical.

Quality control of document text that originates in hardcopy includes use of automated spelling checks, with editing tools for human correction of errors, along with tools and controls for maintaining the dictionary of valid spellings. The dictionary will grow rapidly during early phases of data capture as numerous technical terms and proper names are added. Periodic reviews of the dictionary, along with the careful control of additions, will ensure that few errors are introduced.

Any omissions of material must be explained in a revealing and consistent way. Formulas or other text containing non-roman characters, illustrations, graphics, and tables that cannot be rendered as ASCII text must be identified. Missing portions of text in the original must be so indicated.

As the data base grows, it may be worthwhile to develop software to verify citations in document text against catalog data. Such checks would be useful to detect possible errors in the data and to identify material for inclusion in the LSS data base.

Random checking of the data by staff knowledgeable in the subject matter must be performed on a regular basis. In a data conversion effort for another government organization, such checks revealed the consistent interpretation of a peculiar typographic form of "waves" as "wives". Software can be developed to detect, and perhaps perform context-sensitive filtering of, recurring errors.

2.1.4 Preparation, Transfer and Loading

Following release from the capture quality control process, the document and regulations full text, associated catalog data, and bit-mapped images will require preparation, transfer, and loading into the LSS data base. Differing requirements for access and storage imply that the data will be partitioned into various major data bases such as:

1) A records data base containing the full text, cataloging data, and indices for documents

2) A regulations data base containing the full text, cataloging data, and indices for relevant Federal, State, and local regulations

3) An issues tracking data base that contains the issues information and associated structured indices

4) A commitment tracking data base that contains the commitments information and associated structured indices

5) An image data base that contains the bit-mapped image data for all page images and indices for retrieval of such data.

Since the tracking data can be entered directly into the data bases as described in Sections 2.1.1.1 and 2.1.1.2, no further processing is required.

2.1.4.1 Headers and Full Text

Because of the size of the records full-text data base, it and possibly its associated catalog data are likely to be partitioned into several data base segments. The first step in preparation for storage is to sort the records according to segment. The separate files are then transferred to the computer(s) on which the data base loading will be processed. The loading will store the header data and full text as well as construct or update the structured indices for the header data and the inverted index, which contains pointers to each unique word and its location within the full text. Both the structured and inverted indices should provide quick data access in response to user queries.

For regulations, no segmentation will be required; thus the files prepared during capture can be directly transferred to the loading process, which is identical to records loading.

If full-text searching is implemented <u>via</u> special full-text search hardware then the full-text of the records and regulations will be separately loaded and no inverted index will be built.

If there is more than one node, then duplicate copies of the prepared files will be made and distributed to each node for loading.

2.1.4.2 Images

The bit-mapped images stored during the capture process should include unique accession numbers and page numbers which link them to the catalog data and full-text data bases for records and regulations and support retrieval. The captured images will be transferred to the image storage system for loading.

An alternative method for storing images for display and printing is microform. In that case, the microform is produced during or prior to the capture processing. For the case in which the images are to be available on-line, in preparation for loading into the retrieval storage device, a link tying accession and page number to microform will be constructed.

2.2 Storage of Information in the LSS

The Preliminary Data Scope Analysis concludes that the LSS should support all the information needs of all parties involved in repository licensing, serving as the sole basis for expedited document discovery. The report estimates that this will constitute from 9 to 11 million pages of relevant information in August 1990 (when the LSS is scheduled to be operational), increasing to between 32 to 42 million pages in the next twenty years. NRAC interprets this to mean that all (technologically suitable) material should be in searchable full-text. If this capability is to be provided by software, the associated inverted indices (see Sections 2.1.4.1 and 2.3.2.2) would also be stored by the system.

That report also concluded that for a data base as large and diverse as that envisioned for the LSS to be useful, the material it contains should be extensively, accurately and consistently cataloged. This cataloging information, stored as headers associated with each document in the Records and Regulations Access Subsystems, would also be stored along with their structured indices. Such headers are also likely to be required by the NRAC.

The Preliminary Needs Analysis found that some records may be needed in the form of bit-mapped images while the NRAC is expected to require that images of most pages (in some form) be available through the system. This can require significant additional storage in some designs.

In addition to the storage of standard documents, the LSS is required to store several other types of information. The OCRWM Quality Assurance Plan (DOE, 1986) requires the LSS to be the repository of all Program QA records, which constitute a special class of documents requiring special treatment. Other information, in the form of headers only, must also be stored in the LSS. The statement of work of the LSS Design and Implementation Contract (DOE, 1987) requires the LSS to be able to store indices of the LSS archives. As discussed in Section 1.2, these archives are made up of Document Files and Satellite Files. The Satellite File index should effectively be linked to a Sample Inventory and Management System (SIMS). The NRAC also intends to require the system to store information identifying (but not including the contents of) privileged documents.

Although not large in comparison to the amount of material relevant to the Records Access Subsystem, information in the Tracking Subsystems and the Regulations Access Subsystem will also require some storage capacity.

2.2.1 Contents of Tracking Subsystems

The Tracking Subsystem's data consist of both definition information and status data, in the form of ASCII text. Most of the data elements that describe both types are also stored as structured indices. The definition information is relatively static, while the status data is fairly dynamic. Thus the data should be stored on media that can be easily modified.

For both issues and commitments, once a definition has been entered, an on-line record would likely be maintained for the duration of the LSS operational period. The estimates for the amount of storage required are:

Issues - for 5000 issues, investigations, studies, and activities: the total data storage would be about 13 million bytes and the structured indices would be about 10 million bytes.

Commitments - for 3000 commitments/action items: the data storage would be about 9 million bytes. The structured indices would be an additional 4 million bytes.

2.2.2 Contents of Records Access Subsystem

The Records Access Subsystem storage requirements derive from the need to store the full text of documents subject to discovery during licensing hearings and appeals as defined by the NRAC. In addition to the full text of documents, the Records Access Subsystem document headers, images of non-textual pages, structured indices and inverted indices, which support the full-text retrieval of the document text, also contribute to the storage requirements.

2.2.2.1 Standard Documents

Standard documents stored in the Records Access and Regulations Access Subsystems should consist of header records, ASCII text records, and bit-mapped images records for non-textual document pages. Headers only will be stored for special cases, such as non-document information, documents not included in full text, and privileged information as discussed in Section 2.2.2.2 below. In accordance with ANSI/ASME NQA-1 (1983), Quality Assurance Program Requirements, OCRWM Quality Assurance records should be included in the data base to indicate the level and status of required quality assurance processing.

2.2.2.1.1 Headers

Headers, which are manually input from terminal keyboards, contain cataloging information consisting of such fields as author, corporate affiliation, document title, number of pages, etc. As indicated in the Preliminary Needs Analysis, the following field descriptions were found desirable by a majority of users:

- o Originating and Recipient Organization(s)
- o Date the Document Was Created
- o Author(s)
- o Document Type
- o Baseline Data Flag
- o Submitter Reference Number

Based on the document counts estimated in the Preliminary Data Scope Analysis (see Appendix B), the amount of storage space required to accommodate the header data for all documents and non-documents is:

Time	Storage in megabytes		
	Data	Structured Indices	Total
1990	1,500 - 3,800	300 - 800	1,800 - 4,600
1994	4,600 - 15,000	900 - 3,000	5,500 - 18,000
1998	6,400 - 21,000	1,300 - 4,200	7,700 - 25,000

2.2.2.1.2 Text

The full text of a document should be entered and stored in the LSS and be available for full-text search and retrieval by software or hardware. As a part of the document selection process, the user should be able to browse or read all or parts of the retrieved document.

In order for all the words in documents to be searched by software, the text must be indexed. Software full-text search programs include the tools to accomplish this process; thus it is a relatively automated process and does not require skilled information management personnel. The resulting index, sometimes referred to as an inverted index, contains a sorted list of all words in the documents (except specified common words such as a, an, the, etc.) and a pointer to the storage location(s), i.e. disk volume, file, record, of the words in the documents. The size of the inverted index is a function of the program which is used for the indexing, but it can vary from 50% to 200% of the size of the original ASCII text file.

Full-text inversion, although not labor intensive, requires major computer resources and time to process large files. The files may require segmentation, although this may be invisible to the user.

Based on the page counts estimated in the Preliminary Data Scope Analysis (see Appendix B), the amount of storage space required to accommodate the full text ASCII and its inverted indices is:

Time	Storage in megabytes		
	ASCII Text	Inverted Indices	Total
1990	9,000 - 15,000	8,000 - 17,000	17,000 - 32,000
1994	29,000 - 57,000	25,000 - 69,000	54,000 - 126,000
1998	39,000 - 80,000	35,000 - 95,000	74,000 - 175,000

2.2.2.1.3 Images

Images of documents should be stored in the LSS for the distribution of copies to users on request. Images may be stored in the following forms:

1) Hardcopy (paper)
2) Microform (microfilm, microfiche)
3) Electronic format (compressed bit-mapped images).

The form of image storage depends on two factors: (1) the requirements for the time of receipt of hardcopy as an output of LSS, and (2) the location of the user who makes the request for the hardcopy of a document. The Preliminary Needs Analysis discovered that on the average managers would be willing to wait 3 hours on the average for documents of 100 pages or less. Technical users would wait 7 hours and regulatory staff over 22 hours for the same materials. For hardcopy more than 100 pages, most users would wait overnight.

If a maximum waiting time of 3 hours for a hardcopy of a document of 100 pages or less is considered an LSS requirement, the location of the user who made the request must be considered. If the requesting user's location is local to the LSS node that contains the image, then any of the three possible image forms is acceptable. The hardcopy could be produced by photocopy of the paper image, blow-back from the microform image, or laser printout from the bit-mapped image and hand delivered to the user. However, if the requesting user's location is remote (in another city or state) to the LSS node which contains the image, then electronic transfer (_via_ telecommunications) of the bit-mapped image is the only option available to meet the receipt time requirements.

For the overnight delivery of hardcopies of more than 100 pages, it is possible to meet the requirement with any of the three possible forms of image storage.

Based on the page counts estimated in the Preliminary Data Scope Analysis (see Appendix B), and assuming all images are in electronic form, the amount of storage required to accommodate the bit-mapped images is:

Time	Storage in gigabytes
	Bit-mapped images
1990	170 - 260
1994	540 - 1,000
1998	750 - 1,400

2.2.2.2 Special Cases

The storage and access to some LSS information is considered as special cases. This information includes documents and non-documents in the LSS Archives and privileged information or documents that will not be stored in the LSS.

The LSS Archives should be an integral part of the LSS and should be places to physically deposit those products of the OCRWM program that must be retained and accessible throughout the life of the OCRWM program. The LSS Archives should be able to store documents (including microform) and other physical forms of materials that might include samples of soil, rock, water, plant, etc., and data from tests or explorations that are in non-reproducible form, or in electronic form which, in its unprocessed state, is not useful for inclusion in the on-line portion of the LSS.

2.2.2.2.1 Satellite File Indices (SIMS)

Satellite Files are defined as that portion of the LSS Archives which contains (in boxes, cabinets, shelves) physical forms of materials (such as core) and non-reproducible test or exploration data (such as recorder charts). The on-line portion of the LSS data base will contain only header information concerning the material, indicating where in the Sample Inventory Management System (SIMS) further detail can be found. The SIMS will provide, for each unique item, a description of the item and the physical location of the item for access purposes.

2.2.2.2.2 Document Archive Indices

Document Archives are defined as that portion of the LSS Archives that contains the images (hardcopy, microform, or electronic form) of all LSS documents, with the exception of privileged information documents. As stated in Section 2.2.2.1.1 above, headers for all documents stored in full text in the LSS will be included in the LSS data base. Also headers only will be included in the LSS data base for all document images contained in the Document Archives that have not been stored in full text in the LSS data base (for what ever reasons). These "header only" entries will contain the same information as full-text document headers and in addition will contain the location within the Archive of the document image.

2.2.2.2.3 Privileged Information

Information relevant to the repository licensing that is deemed privileged by any party to the process will be identified in the LSS through the entry of a header with a limited number of fields, which includes a brief description of the document and the privilege being claimed. No entry of text or image will be made until and unless the claim of privilege is overruled. The document will then receive expedited process for entry into the LSS data base in searchable full text.

2.2.2.2.4 OCRWM QA Records

Quality records play a particularly important role in the OCRWM program and the repository licensing process. As such, they require special identification. Quality assurance audits, quality policy manuals, implementing procedures, activity plans, inspection records and test results are examples of quality related records. Codes in the header will indicate that records are quality related. Another header field will indicate the quality level associated with each OCRWM program-generated document.

Often quality records are in sets or packages, such as all of the records related to one QA audit. To associate or tie such a set together, a package header will be used. It will contain the package name, date, source, and identifier. The package identifier will also be in the header of each record in the set. Structured indexes will be created for the package identifier in both the package header and the record header. Other fields in the package header will also have structured indices.

2.2.3 Contents of Regulations Access Subsystem

The Regulations Access Subsystem, like the Records Access Subsystem, has headers, full text, bit-mapped images and the associated indices for each which are stored to support on-line access. A unique aspect of the regulations data base is the requirement to maintain the latest contiguous version of the document even if the regulation is amended in parts.

The amount of storage required, based on the Preliminary Data Scope Analysis (see Appendix B), is:

Time	Header storage in megabytes		
	Headers	Structured Indices	Total
1990	1.1 - 2.1	0.2 - 0.4	1.3 - 2.5
1994	1.3 - 2.6	0.3 - 0.5	1.6 - 3.1
1998	1.6 - 3.2	0.3 - 0.6	1.9 - 3.8

Time	Full Text storage in megabytes		
	ASCII Text	Inverted Indices	Total
1990	43	39 - 52	82 - 95
1994	54	49 - 65	100 - 120
1998	65	58 - 78	120 - 140

Time	Image storage in megabytes
	Bit-mapped Images
1990	350 - 500
1994	440 - 630
1998	530 - 750

2.2.4 LSS Storage Topology

The key objectives for the LSS are the capability for:

1) Full-text storage/retrieval of a large number of documents
2) Rapid, full-text search
3) Full-text access from diverse geographic locations
4) Hardcopy production at terminal locations.

Each objective has a direct bearing on the LSS storage topology, that is, where the LSS data base is physically stored. Similarly, the LSS storage topology has a direct bearing on the total life-cycle cost of the LSS, which includes the development, user/operator training, operations, configuration management, facility preparation, equipment maintenance, telecommunication usage, and data base maintenance.

If it were not for the objective of full-text access and responsive hardcopy production at user terminals at dispersed geographic locations, then a single node (central computer system and data base storage location) would be the most cost effective LSS storage topology. The Preliminary Needs Analysis indicates that there are three major and ten other potential locations across the country for LSS users. The Preliminary Needs Analysis further suggests that in the 1990 to 1992 time frame, the distribution of usage among these locations is approximately:

Washington, DC Area	50%
Las Vegas, NV	30%
San Antonio, TX	10%
Other Locations	10%

Based upon this distribution of potential LSS users, and with 80% of the users located in either Washington, DC or Las Vegas, NV, it is clear that no more than two LSS storage topologies need to be considered as follows:

1) A single-node configuration (central computer system and data base storage) in the Washington, DC or Las Vegas area, and

2) A two-node configuration (two computer systems and distributed data base storage) in both the Washington, DC and Las Vegas, NV locations.

2.2.4.1 Workstations

Because LSS user workstations may differ widely in type and configuration and because of the quality assurance and data base integrity requirements of LSS data, it is nearly operationally impossible to store the LSS data at the workstations. However, the Preliminary Needs Analysis states that about half of the potential users interviewed stated requirements for downloading LSS information to other systems (which include PC-based workstations). The average and peak downloads generally ranged from two to three pages to 20 to 30 pages. However, some regulatory staff indicated peak download requirements in excess of 1,000 pages. Based on these requirements, workstations requesting the download of LSS data would require from 9,000 to 3 million bytes of disk storage.

2.2.4.2 Node(s)

The system design options for the LSS storage topology (i.e. the number and locations of LSS nodes) must address not only the computational, storage, and input/output rates required to meet the functional and operational requirements of the LSS, but additionally must consider the total life-cycle cost of the LSS. Experience has shown that the operational cost of a system, such as the LSS, far exceeds the initial developmental cost, and therefore the recurring operational cost will be a determining factor in the optimal design of the system.

Based upon the geographic distribution and temporal distribution of usage demand contained in the Preliminary Needs Analysis, a central one-node configuration located in the Washington, DC area or Las Vegas and a distributed two-node configuration located in the Washington DC and Las Vegas, NV areas need to be considered as LSS system design options. These two options provide access (via direct terminal connection or low cost Local Area Networks) to the LSS for approximately 80% of the users and support 80% of the usage demand.

2.3 Access to LSS Information

The Preliminary Needs Analysis concluded that structure index searching via detailed and extensive headers should be available, involving subject terms and keywords assigned with the aid of a controlled vocabulary. Full-text search capability on both document text and headers should be also available. These methods should be supported by extensive search aids. Both access methodologies are expected to be required by the NRAC. Further, the system should be easy to use with a minimum of training, should contain built-in help functions, and should provide assistance when needed, either through an expert system or on-call assistance.

2.3.1 Access to Information in the LSS Tracking Subsystems

Each of the user groups identified in the Preliminary Needs Analysis report will also be users of the Tracking Subsystems. Some of the uses will be to:

1) Track licensing and regulatory related issues for the geologic repository program
2) Follow the progress on satisfaction of the Site Characterization Plan (SCP) issues
3) Follow progress on the completion of commitments and action items.

Both the Issues Tracking and the Commitments Tracking Subsystems consist of records, similar to the LSS header, which contain fields such as titles, identification codes, keywords, contact persons and organizations.

All of these fields can be used to search and to specify which records are to be contained in requested reports. The ITS will also have a field which is the full text of the issue. This text in this field will be searchable in a matter similar to the full-text searching of LSS document text.

Both prompting menus and fill-in-the-blanks screens will be available for specifying search criteria.

2.3.2 Access to Information in LSS Records and Regulations Data Bases

Because of the very large and complex data base in the Records and Regulations Access Subsystems, the capability of efficiently and accurately identifying and accessing the information sought is crucial to the usefulness of the system. The Preliminary Needs Analysis has shown that LSS users will require access to LSS data through both structured index searching (on header fields) and through full-text searching. These access capabilities are also explicitly required by the Negotiated Rulemaking Advisory Committee.

The ultimate purpose of any computer-based text storage system is, in response to a specific query, to retrieve the largest possible fraction of documents that are appropriate to the query (high recall) while retrieving as few as possible documents that are inappropriate (high precision), all within the time and complexity constraints of the user. Rather than attempting to identify all "hits" and include no "misses" in a single pass, a very large system such as the LSS uses the concept of set refinement. Here, a retrieved set of results, which has been conservatively defined by an initial query to contain a very large number of possible hits (hopefully all) but also contains much irrelevant material, is iteratively refined by applying additional constraints until most (hopefully all) irrelevant material has been removed with little (hopefully no) relevant material lost. The user is interactively guided during the refinement process by being given the number of hits in the current retrieval set and by being able to examine members of the set to assess the set's degree of relevance (and therefore the degree to which the query has been completed). The results of multiple searches can be logically combined to create hybrid sets (sets that contain the results of multiple queries). Both the results of the queries and the queries themselves can be stored by the user, for later use. Thus, complex routine queries can be easily re-run.

In the LSS concept presented here, the refinement of the retrieval set of a query can be achieved through structured-index searching (Section 2.3.2.1) or full-text searching (Section 2.3.2.2) or a combination of both search techniques.

2.3.2.1 Structured-Index Searching

Structured-index searching, the conventional method used by data base management software to access data, searches indices constructed to support the specific type of queries. These indices are constructed on one field or a combination of fields in the header of a document, such as author, date published or keyword. They can be considered as a surrogate data base that has been specially sorted to support rapid retrieval and response to a specific type of anticipated query on a specific field (or set of fields) in the header. (See Section 2.1.2.3 for the method by which header information is obtained. See Section 2.2.2.1.1 for header and index storage size requirements.)

The completeness of the header design and content (obtained by reliably anticipating the header information which will be needed by users and the ways in which they will search that information) is key to efficient structured-index searching. Significant effort in these areas will be required during detailed LSS design, since structured-index searching takes advantage of a certain amount of pre-processing of the query, and since full-text searching (discussed in the following section) can yield an unmanageably large (and therefore not useful) number of hits when operating on a very large data base. One approach would be to limit the initial steps in the identification and refinement of a retrieval set to structured index searching. Once the set has been reduced to a manageable size (and full-text searches are useful) both methods may be used to resolve the set. The threshold for when both access methods become available must be designed based on the expected response time of a full-text search request.

The ability to formulate Boolean logic expressions among header fields is necessary for useful full-text searching. These expressions need not be directly created by the user, but in order to maximize ease of use, they may be constructed by the system from information prompted from the user. Structured-index searching also requires the creation and storage of index files. The tradeoff between index storage requirements and availability of indices must be examined during the detailed design, but structured indices generally require much less storage than full-text search inverted indices (see next section) because they are based on the less voluminous header information, rather than on the entire text of the documents.

2.3.2.2 Full-Text Searching

Full-text searching is a computerized text processing technique that locates the occurrence of specific words (or groups of words) within text files. In addition to simply finding words, full-text search systems can locate strings in logical or physical relation to each other in a file.

The logical relationships can be specified by Boolean logic expressions when formulating the search condition (e.g. "Find places in the text where 'hot' and 'cold' occur within the same paragraph"). These expressions need not be directly created by the user, but in order to maximize ease of use, they may be constructed by the system from information prompted from the user.

Full-text search capabilities may be implemented in a system such as LSS via either software or hardware techniques. Software techniques create a special file of sorted pointers (called an inverted index) that contain each word in the data base (excluding a set of pre-defined "stop" or trivial words) and their location in the text. In responding to a search, the full-text software uses these locations to respond to the user, performing logical operations on information in the inverted index to respond to queries containing logical constraints. Response time in locating a word can therefore be appreciably shorter than finding the response set to a complex Boolean search. (See Section 2.1.4.1 for a discussion of the creation of the inverted index and Section 2.2.2.1.2 for inverted index storage requirements).

Hardware implementations of full-text searching do not use inverted indices, but rather operate directly on the text data base. During a query, the portion of the data base being searched is streamed past a series of comparators that evaluates it with respect to the query. Matches are recorded and reported to the user. Hardware implementations are also able to perform a Boolean search, and are similarly slowed by logical evaluation over simple location queries.

Hardware full-text searching requires specialized hardware, is a somewhat newer technology and therefore has associated with it a greater development risk. A detailed evaluation of the relative merits and the selection of either a hardware or a software implementation of the required LSS full-text search capability is required during the detailed design. These alternatives are considered significant variants in the conceptual design presented in this report and will also be evaluated in the forthcoming Benefit-Cost Analysis.

2.4 Output of LSS Information

The Preliminary Needs Analysis established the requirements for information to be provided to the user in two basic forms. First is information that is made available to the user in electronic form through an interactive workstation including video displays, printers, local storage, and other capabilities. Second, users require a hardcopy of the document that is faithful in appearance to the original. Perceived response time requirements for both forms are given.

The Preliminary Needs Analysis established several requirements for the ability to interactively query the LSS for information and for the supply of information to the user to be in a form that can best aid the users in performing their job functions. These include, for example, a list of documents that match a query, the text of a specific document, summary reports, and copies of documents representing the "original" format. Most of these requirements will be met through the design of the user workstation, the general term used for the user/machine interface with the LSS.

2.4.1 LSS Output

2.4.1.1 Workstations

The ability to communicate effectively with the LSS from a local workstation, incorporating the major functions that have been identified in the Preliminary Needs Analysis, requires the minimum hardware configuration of a basic personal computer. This level of capability has therefore been defined as the basic or Level 1 workstation configuration. The high availability, relatively low cost, and the fact that many users may already have this hardware in use are factors which lead to this conclusion.

The functional capability to view electronic (bit-mapped) images on the screen at the workstation requires enhanced capabilities to decompress image files, display high resolution images on a screen, temporarily store images, print images, and provide higher throughput communications with the host. Since this capability was not identified as necessary to all potential users, the enhanced workstation (designated Level 2) is not required for LSS access. The summary descriptions of these workstations and their associated capabilities are compared in Table 1.

Since the user will require interactive communication with the LSS, an alphanumeric video screen with keyboard is assumed to be the basic communication device, due to its wide acceptance in the computer industry. This does not rule out the assistance of other devices such as a "mouse", however, since the LSS is a textual data base, a keyboard is a necessity. This will meet most of the functional requirements of entering queries, receiving responses, and viewing document text that are required to determine the relevance of a document to the user's needs. A minimum definition of a keyboard for the workstation would be a standard typewriter QWERTY key arrangement with upper and lower case, shift, return or enter key, and cursor control arrows.

TABLE 1. COMPARISON OF LEVEL 1 AND LEVEL 2 USER WORKSTATIONS

	Level 1	Level 2
Description	Personal Computer Floppy disk and/or hard disk* Monochrome text monitor Keyboard Dot-Matrix printer* Modem or Network connection	Personal Computer Hard disk High-resolution full-page monitor Keyboard Laser printer* Network connection Decompression board Mouse*
Functions	E-Mail: send and receive, upload/download* Data: queries and reports, view ASCII text, download data*, request hardcopy	E-Mail: send and receive, upload/download Data: same as Level 1, plus ability to view, download and print images

* Optional

For Level 1 workstations (without the capability of displaying electronic images), the minimum video screen is an alphanumeric (text mode) monochrome display capable of providing an 80 column by 25 line display. For the Level 2 workstations (with image and full page ASCII text display capability), the preferred system would be a high resolution graphics display in a landscape format, capable of displaying a text display along side of an image display or two image displays simultaneously. Since the image will be stored in a compressed form, the decompression board (assuming decompression is part of the workstation function) must be compatible with the compression techniques used.

Print capability at a workstation will be similarly dependent on the image display (and thus, image reception) capability. For the text-only workstation, a local print capability of a dot matrix (or impact) printer will be sufficient. For those workstations with image display capability, a laser printer would be required, although it would be sufficient to share a printer among several terminals located in the same work area.

The capability to download data from the primary LSS data storage to the LSS local workstation implies an intelligent terminal with capabilities similar to a personal computer, i.e. the workstation must have local storage sufficient to receive the data and file transfer capability to put the data into a transportable format (electronic re-transmission or floppy disk for example) or be able to process the data locally (word processing, for example). This function therefore could be met by defining the LSS workstation to have the basic capabilities of a personal computer. Such a definition is not inconsistent with the keyboard, display, and print capabilities noted above.

2.4.1.2 Printers

The LSS user requires that a hardcopy of selected documents be made available to them for personal use. If the document has existed in paper form (i.e. not originally submitted in electronic form), then the hardcopy should be made from an image of the original paper form. This function can be accomplished in one of two ways. Each (Level 2) workstation could include a laser printer, and the electronic images of the document could be transmitted from the LSS image storage system to the workstation for printing. Alternatively, the hardcopy could be made at the locations where the image is stored and transmitted to the user by express delivery. Considering the difficulties presented by the transmission of images to all workstations (see Section 2.5.3), some combination of these alternatives appears reasonable (e.g., the use of express delivery for large documents and for all documents to locations without image/laser printer capabilities).

2.4.2 Types of LSS Output

2.4.2.1 ASCII

It is clear that for an alternative workstation to permit user queries to a remote textual data base, a minimum capability of transmitting and receiving ASCII format data is required. With the additional requirement for downloading of ASCII files, a workstation based on a personal computer would provide these capabilities.

2.4.2.2 Image

The requirement for electronic images of documents was identified in the Preliminary Needs Analysis, at least for those pages of information that are not capable of being translated into ASCII such as maps, figures and other such graphic information. Similarly, if the format of the ASCII version of a page is sufficiently different from the original (because of the displacement of footnotes or page numbers, for example) for a comparison to be useful, an image of the text pages can also be needed. Since the requirement has been expressed by potential users as well as by the NRAC for storage of images by the LSS (to produce hardcopy), the question remains only to determine the method for making these images available at the workstation. Alternatives are:

1) Store the image in electronic form (or a form readily converted to electronic form) and transmit the image to the workstation as requested.

2) Store the image in electronic form on a reproducible media such as CD-ROM, distribute the media, and provide the capability at the workstation to locate and display the image.

3) Store the image on a reproducible media such as microform and distribute the media along with a location index and a reader.

In any event, it is clear that not all users of the LSS require that images be available to them at their workstation, thus enabling the definition of workstations with and without image capability, and providing only a fraction of the workstations at a facility with image capability.

With the anticipated data base of the LSS records extending to tens of millions of pages, the distribution and local storage of all images at the workstation either in electronic form or microform, becomes a major operational problem. For example, at current densities of storage the LSS administrator would be distributing 2 to 3 CD-ROM disks per day to each user. Therefore, local distribution of images is viable only if those images are limited to pages or portions of pages that are unsuitable for conversion into ASCII format. This solution presents the difficult configuration management problem of separating images of graphics from the parent document and ensuring that the dislocated pages are correctly filed and distributed.

The magnitude of the problem is a function of the percentage of documents which contain graphics and the percentage of pages within those documents which are graphic; neither of these two percentages have yet been defined to an accuracy that one can assess the magnitude of the problem in detail. The conservative assumption for the conceptual design was therefore made that images at all of the workstation will be provided over a communications link from a central image storage location. Since all images will be available at this location, it is not necessary to differentiate between images of graphics and images of text pages. An alternative is also considered where there would be no images provided directly to workstations. The actual design will probably fall between these extremes, i.e. images may be provided to some but not to all workstations.

2.5 Communication of LSS Information

The Preliminary Needs Analysis estimates that the LSS will need to support access to the LSS from multiple sites, in over a dozen cities throughout the U.S. The majority of LSS workstations are expected to be in the Washington, DC and Las Vegas areas.

In addition to the explicit communications requirements associated with connecting workstations to the LSS, several implicit communications requirements also arise from potential system designs.

2.5.1 Node to Node Communication

If the LSS computer system is not centralized at a single site, there would be a requirement for communication between all of the LSS computer processing subsystems (or nodes). There is a wide range of possibilities for communicating information from node to node. This range includes:

1) A dedicated, real-time communication facility
2) A shared, batch communication facility
3) An overnight, courier delivery service.

The main determinants on which communication environment would best meet the LSS requirements are the time-sensitivity of LSS information, geographic proximity of LSS processing nodes, and the LSS data base structure and requirements.

If either of the electronic communication environments are required, each node site would require some communication equipment. Each node would have to dedicate physical ports on their data processing hardware for communication. These ports would need to be locally connected to a communication signaling device (such as a modem or multiplexer). The signaling devices require a communication circuit (such as a telephone line) which extends between the two node locations.

In order to reliably communicate electronic information, each LSS node must be capable of supporting a defined set of rules, i.e. a protocol. The protocol establishes the vehicle for the flow of information between LSS node sites. Various protocols exist for various data processing applications. The appropriate protocol must be supported by the various vendors' hardware and software that comprise an LSS node site.

2.5.2 Capture Station to Storage Communications

If the LSS capture station is not physically co-located with the LSS data storage facility, then there will be a need to transmit the captured data to the storage facility. Like the node-to-node communication environment, the capture-to-storage process also has three alternate configurations:

1) A dedicated, real-time communication facility
2) A shared, batch communication facility
3) An overnight, courier delivery service.

If either of the electronic alternatives are chosen, there are certain generic requirements for communication. The capture workstation must have an interim data storage device that serves as a buffer and back-up mechanism. This data storage device could either be connected to a communication device directly or it could transmit its information through a communication device associated with the capture station. These local communication requirements (capture device to interim storage device) would be dictated by the vendor of the capture device.

To move the information from the interim to the permanent storage location would require a physical port on each device. These ports would have to be connected to a communication signaling device (such as a modem or multiplexer). The signaling devices require a communication circuit (such as a telephone line) that extends between each capture node to the data processing node. The speed and type of these connections depend on the time-sensitivity of the data.

Overnight mail of high density storage media (such as magnetic tape for ASCII files or optical disks for images) can well be a viable communications mode for sending information from a capture station that is not near an LSS node. This requires compatible capability for reading and writing such disks at both systems.

2.5.3 Node to Workstation Communications

The function of this communication is to provide reliable access to the LSS data base with the added features of downloading data to local workstations and possibly displaying or outputting bit-mapped images locally. Because of the concentration of LSS users in three major geographic areas (Washington, DC, Las Vegas, NV, San Antonio, TX, according to the Preliminary Needs Analysis) the communication requirements should be modeled to support intensive communication from these sites to the processing or host site or sites.

In an effort to pool LSS resources at the seven most concentrated usage locations (White Flint, Forrestal, Weston, M&O Contractor, San Antonio, Las Vegas and Carson City), local area networks (LAN) are recommended. LANs allow users to share common equipment such as laser printers, personal computer (PC) storage devices and communication facilities. Each PC or associated device at each site would require access to the LAN. The access usually entails a physical communication port (or card) and a communication circuit (or cable). The LAN would benefit the user most by having bridged access to the remote LSS host location. This form of remote connectivity allows a LAN user to request large LSS files to be printed or stored at any device attached to their LAN. This offloading of data from the PC maximizes the communication between sites, provides timely/efficient print and storage functions and minimizes the investments in individual workstations while enhancing the capabilities of the total LAN user group.

Once LANs are established at the major usage locations, the shared communication circuit (or telephone line) to the host sites would be sized according to the anticipated traffic of the entire LAN user group. The rate or line speed of the communication circuit is variable and could be sized according to the unique demands of each site. If LSS users and a capture station were co-located, and the capture station was electronically connected to an LSS node, then each application (access and capture) could share a common communication facility to the host site.

For LSS users not located at the seven major usage locations, basic workstation access needs to provide terminal to host communication. Each workstation would need access to the LSS data base and some workstations would need the capability to download files or view images locally. Each workstation would require a physical communication port, communication device, communication software and communication circuit. The specific type of each of these four functions would be workstation dependent. The major variant is the rate at which communication will occur which provides the necessary response time for the desired application (such as viewing images). The most reliable way to provide terminal access to an LSS user would be to extend a dedicated communication circuit from the remote location (e.g. Lawrence Livermore National Laboratory) to the nearest LSS LAN site (e.g. Las Vegas, NV). This dedicated circuit would allow the remote site to share the larger, more economic communication circuit between the LAN site and the host site. For users only requiring limited access to the LSS data base (such as those needing ASCII output only), a dial-up communication circuit at a reasonably high data rate (such as 9.6 Kbps) should suffice most applications.

2.6 Electronic Mail

The Preliminary Needs Analysis found that an electronic mail capability within LSS may be needed. It appears that the NRAC will find that the LSS should be able to accept input from such a system. When this feature is used to electronically file briefs, for example, the submitter should be able to certify that the material captured by the system conforms with what was sent.

2.6.1 Conventional E-Mail Functions

The following LSS functional capabilities can be met with currently existing electronic mail software and therefore require little or no customization.

2.6.1.1 Uploading Files from a Workstation

Since many of the hearing related documents will be multipage messages, it will be more efficient to provide the capability of uploading message text files to the LSS which can then be transferred to the E-mail function (rather than requiring the message to be keyed in on-line). This will allow the user to prepare the text of the message on office word processing equipment "off-line" to the LSS. For consistency, the format for acceptance of the message file into the LSS E-mail system should be the same as the format for word processing files submitted for the LSS records.

2.6.1.2 Message Transmission and Receipt

The filing of hearing related documents will require certain common E-mail functions to be available. Since messages will normally be sent to multiple parties, the system should provide for the ability to send messages to a distribution list. To assure completion of the process, both delivery receipt (notification to sender that the message is in the receiver's "mailbox") and read receipt (notification to sender that someone with the receiver's access code has read the message) functions are required. To provide a more positive indication, a receipt acknowledgment message capability initiated by the receiver to the sender may be provided. Date and time "stamping" of the message will automatically be provided so that the receiver will know how long the message has been available in his mailbox.

2.6.1.3 Privacy and Authenticity

The E-mail system (as well as the LSS in general) can be provided with password security access and identification which will provide privacy on the contents of a mailbox as well as an authentication of the sender's identity.

2.6.2 Special LSS E-Mail Functions

Hearing related documents such as motions and pleadings which are sent out on the E-mail system must be sent as well to the LSS records data base for input. While inclusion of the document could be made automatically (assuming the E-mail function is integral to the LSS), for reasons of quality assurance and management it is preferable to designate one of the capture workstations as an "LSS mailbox" to receive copies of the messages. These can then be processed through the cataloging and capture process in the same manner as hardcopy documents or word processing files. The submitter can be allowed to verify the correctness of the captured file and certify it if required.

2.7 <u>LSS Management Functions</u>

The need for a rigorous LSS quality control process for both data in the LSS and the system itself is concluded in the Preliminary Needs Analysis and is expected to be required by the NRAC. Performance and usage monitoring (required by the statement of work of the LSS Design and Implementation contract, DOE, 1987) is also identified as necessary for providing the data for system optimization.

2.7.1 Performance and Usage Monitoring

To obtain the best performance possible, the LSS includes hardware and software which enable the LSS administrator to monitor its performance. The performance information provided will assist the LSS operators in determining what immediate and future adjustments would be needed to improve performance. The following performance monitoring tools will be included in the system:

Telecommunications - equipment and software which measures and records the transmission loads and data errors on each link in the network as well as equipment malfunctions.

Hardware - software that measures and records the input/output and computation loads on all computers, the loads placed on the magnetic and optical disk storage devices, and the amount of storage fragmentation of the data base.

Software - special software will sample and record the data entry loads and the types of user queries being made.

Analysis of the monitoring data obtained will assist in the identification of hardware needing maintenance, planning maintenance schedules, changing communication routings, relocation of cluster hardware and possible software or data base architecture changes to better support the types of queries being experienced.

2.7.2 LSS System Administration

2.7.2.1 LSS Quality Control

The quality of the LSS will be determined by the accuracy and completeness of its contents, the quality of the cataloging processes, the richness and accuracy of the thesaurus and other retrieval aids, and the reliability of the hardware and software components. To ensure a quality system is maintained, a quality control process should be established that sets standards, develops policies and audits the system's quality. The performance monitoring tools described in Section 2.7.1, the capture quality control tools described in Section 2.1.3, and the configuration management tools described in Section 2.7.2.2 can also assist the quality monitoring function.

2.7.2.2 LSS Configuration Management

Knowing precisely what hardware and software comprise the LSS is required to efficiently manage its operation. To support this need, the LSS design must include configuration management software as tools for use by the LSS administrator. A LSS hardware configuration management software package should include the following:

1) A current listing of the LSS system's hardware configuration, including component vendor and part number

2) A listing of the spare parts inventory of each hardware component

3) A listing of and status of each hardware trouble report, which documents any hardware malfunction requiring maintenance.

Software configuration management tools should be provided or developed for the following:

1) Operating system software generation and update capability to generate and maintain the operating system and utilities to the current, approved version level

2) Software for generating, updating, and controlling the LSS application software consistent with the requirements of the configuration management program

3) A listing of and current status of any software Trouble Reports, which document any software system error that requires maintenance

4) Software for data base management and control including data base backup and recovery and for the periodic generation of data base copies to be stored off-site to prevent loss of the data base due to fire or some other catastrophic occurrence.

2.7.3 Data Base Administration

2.7.3.1 Data Base Maintenance

Data base maintenance for LSS encompasses numerous activities that are critical to system reliability, the usability of the LSS data, and to user confidence in the LSS. Data base maintenance includes activities such as:

1) Appending new data to the LSS
2) Correcting erroneous entries
3) Restructuring to consolidate fragmented space
4) Maintaining the LSS data dictionary containing
   - Descriptions of the data structure and of schema
   - A glossary defining all terms in the controlled vocabulary used for cataloging entries.

The above activities require both supporting software and organizational procedures for executing the actions at routine intervals, defined according to calendar time or specific data base activities (e.g., before and after new data is appended to a data base). The data base administrator should keep a log of all maintenance activities performed on the LSS.

2.7.3.2 Loss Protection

The data base administrator will perform the following activities on a regular basis:

1) Routine backups of the data
2) Execute recovery procedures when required
3) Storage and maintenance of backups, both on-site and off-site
4) Execute various check routines to ensure the physical integrity of the data base.

2.7.3.3 Access Control

LSS users will have one more of the following privileges:

1) Search and retrieve all data in the LSS
2) Search and retrieve all header data in the LSS
3) Search and retrieve data in the Tracking Subsystems
4) Update authorization for the Tracking Subsystems
5) Update authorization to the Records and Regulations Access Subsystems.

User names and passwords will be used to enforce privileges; the system administrator will add and delete passwords for individual users.

3.0 CONCEPT OF OPERATION: DATA CAPTURE AND DATA RETRIEVAL

The preceding section has given a detailed description of the functions needed by LSS users. The various ways in which these functions can be implemented cannot be indiscriminantly combined into a single system. This section presents an operational concept of LSS that integrates these functions effectively to satisfy the requirements identified. The concept of operation presented here is subdivided into the two main operational areas of the LSS: information capture (Section 3.1) and retrieval (Section 3.2). The base conceptual design presented in Section 4.1 represents a possible hardware and software environment that supports the operations described in this section.

3.1 Document Preparation and Data Capture

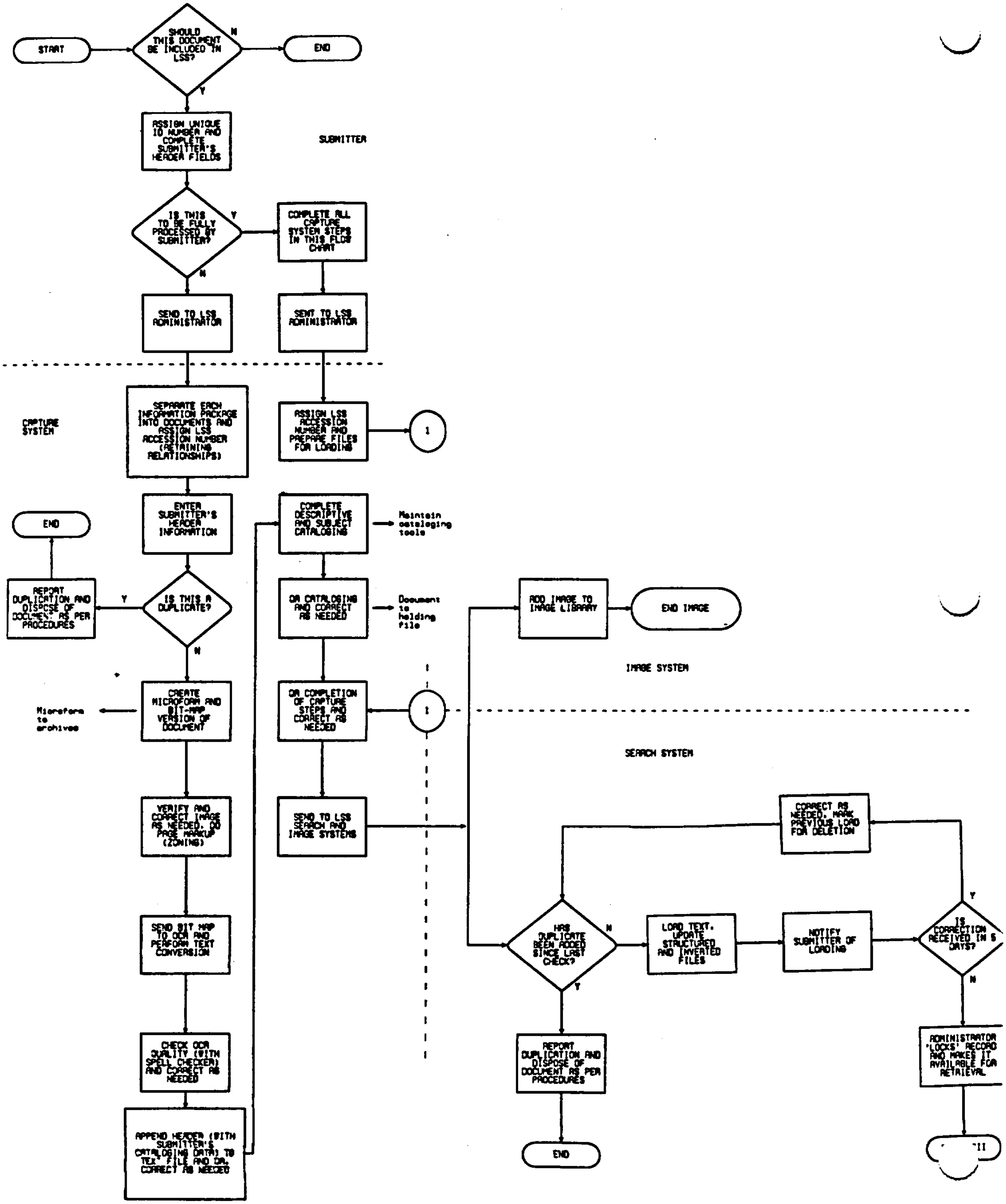
The document flow described here is tied to the Base Conceptual Design. In this concept of operation, items will be processed on a single pass at a capture site. Moreover, this operation is based on the assumption that technology advances anticipated in the next one to two years will be incorporated into the document preparation system. Such advances include the use of hand-held scanners to capture descriptive cataloging data and the development of software to assist in the document cataloging process. In the event that a different document capture or preparation process is adopted (for example, a plan to defer conversion from image to text until some later time), the integrated process described below would change significantly.

The LSS administrator's responsibility for documents begins when the document is received at the capture site. Note that the submitter has the option of fully processing the document. In this case all the capture steps described below, are performed by the submitter. The administrator needs only QA check the material and add the submitted material to the normal capture system information flow. Several stages of document processing must be completed before the document is available to LSS users. Figure 1 depicts this process and associate flows. The processing tasks require both manual and automated support, as described in the following paragraphs.

In this concept of operation, items from authorized submitters arrive at LSS administration as "information packages" that contain one or more related items of information such as printed matter, maps, and computer printouts. Each information package is assigned a unique identification code number, as are the components of the package. This number will also reflect the relationships among the components of the package. The date of receipt and the data supplied by the submitter for each component, along with the information package association, is recorded. The unit of processing from this point on is the individual numbered item.

FIGURE 1

PROCESS FLOW IN THE CONCEPT OF OPERATION FOR LSS DATA CAPTURE



The item is first checked against the data base to determine whether it is a possible duplicate of a previously-received item. If a candidate duplicate is found, a cataloger checks both items to determine whether they are, in fact, duplicates. If they are identical, processing of the newly-received item is terminated. Disposition of the item is recorded in the data base and reported to the submitter. If the item is not a duplicate, processing continues. However, if the item has an important physical or logical relationship to another item in the data base (such as a copy of a document with extensive marginal notes added), the association will be recorded in the catalog data for the item.

Items next proceed to microform conversion and scanning (to create the required microform archive copy and the bit-map copy needed for OCR conversion). The images are verified for skew, inversion, page order and completeness. If a page is missing from a document, a substitute page containing a message to that effect will be inserted. Tools will be available for scanning operators to insert, delete and correct page order.

The next step is to mark up or (zone) the image for text conversion. The mark-up, performed with a mouse, will delineate formulas, figures and graphs not suitable for conversion. For example, if captions are converted to text, captions will be marked so that they can be readily identified after conversion has occurred. In the same manner, abstracts, footnotes, and any other portions of text that need to be explicitly identifiable to text users, will be marked.

If the item is a candidate for automatic text conversion, the image files are dispatched for batch OCR processing. If the item is not suitable for automatic text conversion (e.g., if it is a handwritten document), it is so marked and sent directly to cataloging. The decision to key in the contents of the document is made at this point.

After text conversion the text file is spell checked. Spell checking provides a rapid way to assess the accuracy of the OCR process and to help facilitate correction. Spelling correction will be done by an editor.

The next step is to complete the catalog information. At the cataloging workstation, the cataloger will have access to both the electronic and the hardcopy version of each item. Menus, on-line thesauri, and other manual and automated tools will be available for catalogers to select appropriate values for each header field. The software will check spelling and legal values for catalog data and will verify that all required fields are populated. Any anomalies are marked, and alternative spellings or values will be presented where feasible.

The file is then presented to the quality control (QC) cataloger who will review the correctness of the cataloging data and change any that is incomplete or incorrect. The QC cataloger will also have both electronic and hardcopy versions of the item for these activities. As terms are identified that need to be added to the lists of valid spellings and legal values (e.g., a new term for the thesaurus), the QC cataloger will propose such changes to a technical overview group responsible for approving changes. Once approved, changes will be implemented by a data administrator.

A software verification of the new catalog data is performed. If the file is verified, it is ready for loading. If errors are found, the file is returned to the QC cataloger for correction and re-verification. Since image files, header files, and text files associated with a single item will be added to the LSS data bases as one task, it is also the responsibility of the QC cataloger to ensure that these linked files are consistent, i.e., that they are all identified with the same item. Index creation for text and header data occurs as part of the loading process.

Prior to loading, the system will again be checked for duplicates to ensure that a duplicate item has not been entered during the time the current item was in process. If a candidate duplicate is found, a cataloger checks both items to determine if they are identical. If the items are indeed identical, processing is terminated and disposition of the newly processed item is recorded in the data base and reported to the submitter.

After the files have been loaded, the submitter will be notified. The submitter has five days in which to identify any corrections to the LSS administrator. If corrections are submitted, they will be made in accordance with the appropriate procedures. If no corrections have been received, the LSS administrator locks the record against any changes. The record is then available for full access.

Backups of items in preparation will be performed on a regular basis. Backups of image files will be retained until the file has been backed up as part of the permanent LSS data base. Backups of fully-processed versions of text and header files will be retained until the document has been backed up as part of the permanent LSS data base.

3.2 Data Retrieval Operations

For the LSS to be successful, the operational system that is implemented from the design must meet the needs of each of the user classes identified in the Preliminary Needs Analysis. This section discusses the retrieval operations characteristics of the Base Conceptual Design.

3.2.1 Access

The LSS administrator issues users' names and passwords and assigns privileges such as which LSS subsystems each user may access.

To obtain access for a particular session, a user goes to the location of the nearest workstation. For those in one of the primary LSS user locations (White Flint, NRC; Forrestal, DOE Headquarters; M&O contractor; Las Vegas, NNWSI; Carson City, State of Nevada; or San Antonio, FFRDC), Level 2 workstations, tied directly to the dedicated LSS communication network, will be available in the offices of those whose jobs involve frequent use of the LSS. For users who have less frequent need to use the LSS, areas with both Level 1 and Level 2 workstations, also tied to the dedicated LSS communications network, will be established. Users not in those locations will access the LSS using Level 1 workstations and dial up communications.

The Issues and Commitment Tracking Subsystems will be updated on-line as new issues or commitments are defined and as status information is entered. For the Records and Regulations Access Subsystems, new documents are processed daily as described in Section 3.1, however, access to the recently processed records would be provided on a periodic basis such as every two weeks or once a month. This will ensure that searches and retrievals performed during these periods work with the same data base contents. Upon entry into the subsystem a summary of the newly available data will be displayed.

The LSS would be available for access from 7:00 am to 11:59 p.m. Eastern Standard Time every day except holidays. This will support the vast majority of the access periods identified in the Preliminary Needs Analysis.

3.2.2 Query Operations

Once access been obtained to the LSS and either the Records or Regulations Access Subsystem has been chosen, a user has a choice of four styles of user interfaces for formulating queries:

1) Menus - which allow the selection of specific, predetermined queries from a list

2) Query screens - which provide fill-in-the-blank type of screens to specify which fields should be searched for what terms. Windows which can be used to display the legal values for a field or expand the contents will likely be incorporated.

3) Prompting dialog - in which the search software "converses" with the user via a series of questions to construct a query

4) Query language of the data base software - which provides the most flexibility, but is only for trained, experienced and frequent users, such as intermediaries.

Both the headers and full-text data can be searched using any of the four interface styles. Each query may involve specifying one or more header fields such as subject, title or date; one or more phrases, or two terms within a certain proximity, to be searched in the full text. Results of a query may be combined with earlier query results of either headers or full text using the Boolean operators of OR, AND, and NOT.

The Tracking subsystems' data would be stored as structured data, similar to the headers in the Records and Regulations Access Subsystems, thus full-text searches would not be applicable. Also, menus, query screens and query language of the data base software will be part of the Tracking subsystems.

Upon the completion of each query, the results would be displayed showing the number of instances in which the search criteria have been met and the number of documents involved.

Assistance for the users would be available _via_ an on-line facility as well as through user guides and manuals.

3.2.3 Retrieval Operations

The Base Conceptual Design supports retrieving technical data either as on-line displays or in the form of hardcopy. Special policies and procedures would govern retrieval of non-documents in the archives.

3.2.3.1 Headers

The on-line display of headers would be _via_ formatted screens with each file labeled and its values shown. For some large fields, only a portion of the contents might be displayed. To view all of the contents, the user may scroll through the field's "window", or may "zoom" the field to cover most of the space on the screen.

Headers would be displayed for viewing one at a time, but a user could easily browse through a result set, either forwards or backwards. The user may specify how the result set is to be ordered, before it is viewed. Header information relating to a result set could also be viewed in a tabular listing, with the columns being selected header fields and each row the field contents for headers in the result set.

3.2.3.2 Full Text

In the Records and Regulations Access Subsystems, full text could be displayed. On Level 2 workstations, the text would be viewed in full page mode, formatted similar to its appearance in the original document. Exceptions would be that all text would be displayed in a single size and style, and figures, graphics and formulas would be replaced with a note. To see an exact copy, an image display or hardcopy would be requested. The display of text on Level 1 workstations would be a partial page of up to 24 lines. The ability to scan up and down a page would be provided.

Text, like the headers, could also be browsed from page to page or possibly from document to document. After a result set of documents and their associated full text has been established by a search, the documents in the results could be ordered before browsing. Header fields could be specified to determine the order.

The Base Conceptual Design provides for downloading an ASCII file, containing the text associated with documents in a result set for processing in local workstation mode.

3.2.3.3 Images

The Base Conceptual Design provides for the displaying of Records and Regulations Access Subsystem pages in the form of images on Level 2 workstations. The images associated with a result set could be browsed similar to the manner in which full text is browsed. Images could also be displayed one at a time by specifying the document and page number or by requesting the image associated with the full text ASCII page currently displayed. Level 1 workstations will not have an image display capability.

All LSS users would be able to obtain images in the form of hardcopy. Users at Level 2 workstations would be able to print images at the workstation. There will likely be a per request limit on the number of images that may be printed at the workstation. Large volume and all Level 1 workstation print requests would be routed to the image system for printing and priority shipping.

3.2.4 User Session

In any one use of the LSS, users would combine the query and retrieval operations, described above, to best obtain the information they desire. To estimate the communication, search and retrieval loads that are likely to be placed on the LSS, a series of user session scenarios specifying how search and retrieval activities like those presented in Table 2 will be developed and quantified. An example scenario is presented in Appendix A.

TABLE 2. USER SEARCH & RETRIEVAL ACTIVITIES

Header searches:	Searches on the header or catalog descriptions of documents and non-document material
Return number of hits:	The results displayed are only a report of the number of instances the search criteria were met and of the number of documents in which hits were found
Return header:	The user chooses to have header data displayed (as opposed to a report of the number of hits or to seeing the document text)
Return text:	The user asks to see the text of the document, rather than just the header data.
Image browsing:	The user asks to see page images of the document, in addition to the document text.
Full text browsing:	The universe in which searches are performed includes the full text of documents, rather than just the header or catalog data
Local print requests:	Requests to have material printed at the local site
Printed page images:	Printing image versions of document pages
Print Header information:	Printing header or catalog data
E-mail messages:	Electronic mail messages sent during the session

4.0 CONCEPTUAL DESIGN

Now that a preliminary requirements analysis for the LSS has been completed, consisting of the first two reports in this series, it is appropriate to focus on possible design concepts to implement the requirements identified. No single optimum design exists. Rather, a variety of basic concepts and myriad variations on those concepts can be acceptable. The object of this analysis is to focus on one family of high-level designs, with low development risk (i.e., very likely to meet all the requirements and very likely to be technically achievable). The forthcoming Benefit-Cost Analysis will evaluate the financial feasibility of the family of design developed here.

The conceptual design outlined here is a high-level system design, identifying the major LSS functions, hardware and software subsystems, and subsystem interfaces. It represents the consensus of a team of specialists in information management, software design and development, systems integration, communications, image storage and retrieval, nuclear regulatory development and compliance, and nuclear waste management. The high-level design developed by this team presents the best system (i.e., lowest development risk) to implement the identified LSS requirements and is presented as a first step in evolving a detailed design for the LSS. It provides a basis for further review and discussion.

The basic design concept that has been developed was not selected from an exhaustive list of potentially viable alternatives. Rather, after an examination of previous work on LSS and of similar systems by the design team, a concept perceived to have very low development risk was selected. Other alternatives were not considered in detail. For example, in the selection processes one potentially viable alternative that was considered but not developed involves replicating the LSS data base (in totality, or in part - such as text only, or images only) at each user workstation. This concept requires the regular copying and distribution of new material to maintain and keep consistent over one hundred copies of the LSS data base. Because of the very large size of the data base and the extremely difficult data configuration management problem that such a distributed system would pose, the design team preferred a base design with lower development and operational risk.

The conceptual design outlined here is a concept, having features which can be implemented in a number of viable ways. Not all of the options described are compatible or have the same functionality. The approach taken during this analysis was to develop the base conceptual design and a number of internally consistent variants. Each variant considered introduces a different way to implement a feature of the system and identifies the repercussions of this choice on the other system features and functions. The forthcoming Benefit-Cost Analysis will evaluate the financial feasibility of the family of designs developed here and provide a framework for evaluating the trade-offs between the costs and technical risks versus the amount of functionality offered by each design option.

4.1 Base Conceptual Design

4.1.1 Base Conceptual Design Hardware

The Base Conceptual Design hardware architecture, is illustrated in Figure 2, and consists of several replicated capture systems, the main computing and data storage capacity (called the search system), the image system, communications system and workstations. The following sections describe the hardware and software that constitute this Base Conceptual Design.

4.1.1.1 Capture System

A document capture system consists of a computer system and attached (local) terminals for control of the document capture process, cataloging and correction (if not done off-line). The computer system interfaces and controls, *via* software, the peripheral devices required for the scanning, and optical character recognition processes needed for the capture and quality control of the document images and ASCII text.

Local terminals attached to the computer system, with appropriate software (see Section 4.1.2.1), are used for:

1) Validation and quality control of the digital scanning process, which will capture and store bit-mapped images

2) Validation and quality control of the microform conversion process, which will read the microform media and either produce a hardcopy of the image page or produce an ASCII text record of the image page

3) Validation and quality control of the OCR process, which will produce an ASCII text record of a bit-mapped image

4) Input and validation of documents by re-keying the text of poor quality documents such as those produced from microform or deteriorated hardcopy

5) Input and validation of cataloging information

6) Input, validation, and cataloging of messages from the LSS electronic mail system.

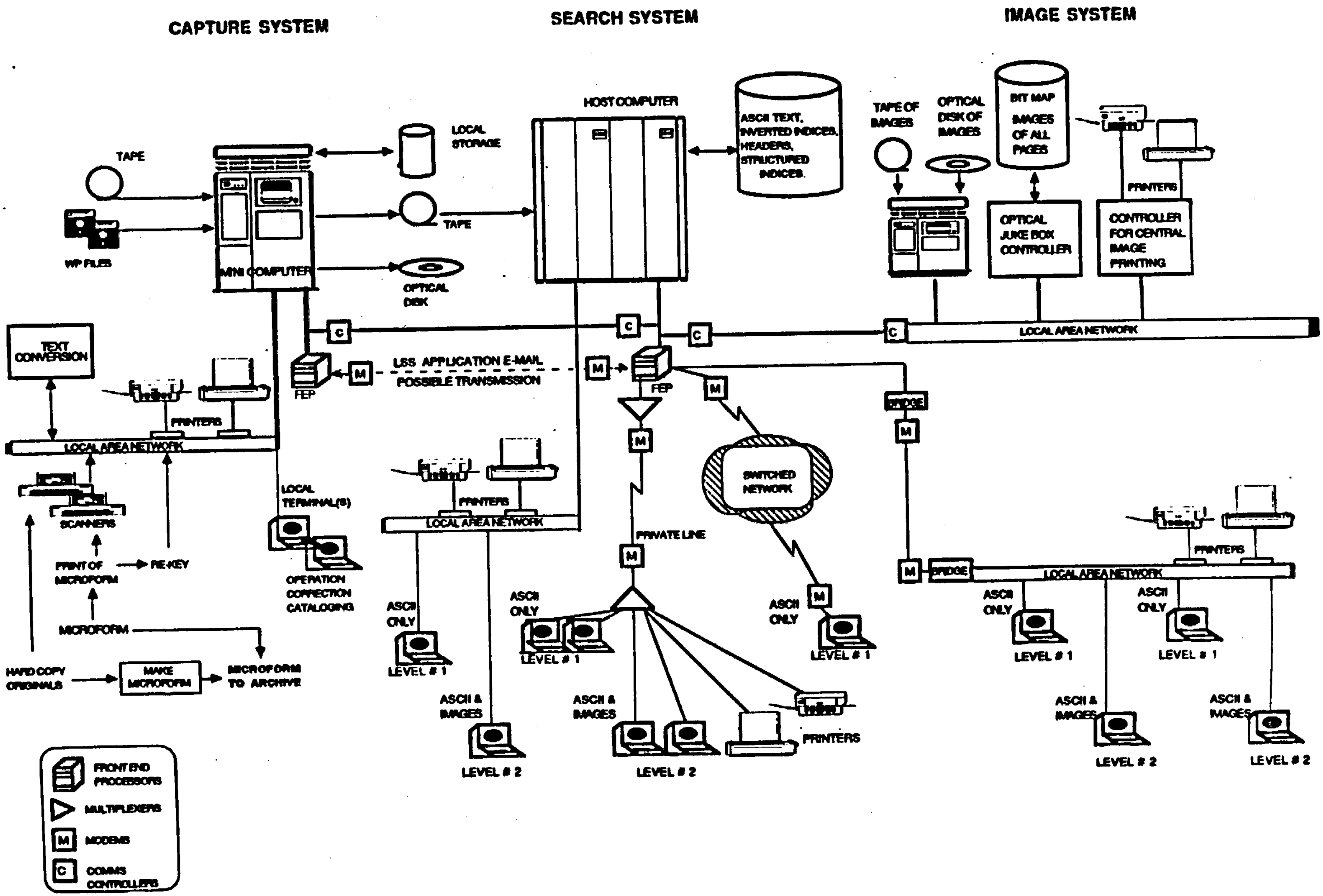
The document capture center's peripheral devices include:

1) A high speed digital scanner

2) A text image recognition and conversion processor

FIGURE 2

BASE CONCEPTUAL DESIGN HARDWARE ARCHITECTURE



3) A microform reader and conversion device for producing hardcopy and/or ASCII text records of microform images

4) Magnetic disk drives for temporary storage

5) Optical disk drives

6) Magnetic tape drives for the storage and distribution of captured data and for input of documents in machine readable format, which have been prepared by word processing software

7) Magnetic floppy diskette drives for input of documents in machine readable format, which have been prepared by word processing software.

The digital scanning process should provide the following minimum capabilities:

1) An automatic document feed capability of 30 to 50 document pages of up to 8.5 x 14 inches in size.

2) Capture each document page as a bit-mapped image with a minimum of 300 dpi (dots per inch) resolution.

3) Perform CCITT Group IV image compression with operator selectable compression ratios.

4) Store the bit-mapped image on magnetic disk coded with an accession number for later retrieval for quality control processing and optical character recognition processing.

As discussed in Section 2.1.2, the capture of material into LSS is to be accomplished at several capture centers, some of which may be co-located with the LSS node.

An electronic image of a document page from hardcopy is captured by a process of feeding document pages into a digital scanning device, checking the resultant image, and entering an accession number from a terminal keyboard for identification of the document. As the document image is captured, the digitized bit-mapped image is displayed on a high resolution page-oriented monitor capable of at least 150 dpi resolution used to verify the quality of the scanned image and allow for re-scanning if the image does not meet the quality control requirements. The image is a replica of the original, including margin notes, signatures, graphics, date stamp, etc. which can not be captured in ASCII form. Images are the only reasonable method of capturing graphic oriented document pages.

Although images are electronic, the characters or words on the page cannot be recognized by a computer until the image is processed by optical character recognition. Electronic images require relatively large amounts of storage, typically 50,000 to 70,000 bytes (compressed mode) per 8 1/2 x 11 inch page, as compared to ASCII at 2500 to 3000 bytes per page. Thus the use of images require high density storage devices such as optical disks.

Documents which have been prepared on a computer by word processing software are already in machine readable format. However, because most full-text retrieval programs require that files be entered in ASCII form and because the variety of word processors systems in use by Government and industry are not standardized for computer communications of ASCII text, some conversion is required. Generally speaking, software tools are available to support this conversion. Most word processor systems provide the option for producing an ASCII file of the document text as an output. This file does not contain the special codes, for formatting and printing, that the word processing document file contains. However, even this ASCII file format is not standardized, and therefore a standard file format must be defined for input to the LSS. The Technical Staff of the NRAC is preparing such a standard, and the LSS should accept and store the ASCII text provided in the approved format.

Archive requirements dictate that all LSS documents must be microformed. The Base Conceptual Design meets this requirement by creating a microform copy of all LSS documents as part of the capture processes.

4.1.1.2 Search System

To the LSS user, the search system _is_ the LSS. It is the single point of contact, supporting all of the LSS functions either directly or transparently through directions issued to the image system and the workstations.

The search system supports a myriad of functions including electronic mail, issues and commitment tracking, search and retrieval services for all document types, storage of all the regulations, header, and full text including their associated structured and inverted indices, update of the data bases as new documents are received from the capture system, downloading of ASCII text to the workstations, and routing of commands to the image system for off-line printing or on-line display of document images. All of these functions must be performed within reasonable response requirements such as those listed on Table 1, of the Preliminary Needs Analysis.

The base conceptual design hardware architecture has a single search system, co-located with the image system. The search system receives data from the capture system(s), which will or will not be co-located with the search system. The search system then updates the LSS data bases, including structured and inverted indices. LSS users are supported by two levels of intelligent workstations connected _via_ the communications system.

The search system can consist of a large mainframe or a tightly coupled cluster of super-minicomputers. Either configuration will provide the computational, disk storage, and input/output transfer rates required. The system will be sized to handle the peak loads associated with a LSS user community of 225 to 475 people at the peak point in the licensing process, according to the Preliminary Needs Analysis. Disk requirements for the search system, estimated in Section 2.2.2, include all of the document types plus structured and inverted indices. The capacity and number of input/output channels are critical for a number of users and the size of the data base that the search system must support. The search system will be designed for incremental growth and for technology insertion of faster processors, larger capacity disks, and higher input/output transfer rates. All configuration changes will be transparent to the users and the software.

Provisions will be made for additional central processors and disk drives beyond those necessary to support the on-line use of the system. These extra computer resources will enable the system operators to:

1) Update data bases 24 hours per day. When a data base is to be updated, it will be copied to the extra disks, updated by the extra processor(s) and the system switched to the updated copy. This ability will be crucial during the initial operation of the system when large numbers of documents are entered every day.

2) Allow continuous backup of the system without affecting on-line response. This need for backup is also driven by the high rate of change in the data bases.

3) Fully develop and test new system and applications software releases prior to going on-line with the LSS users.

4) Provide backup for failed components and to deal with unexpected surges in demand.

Beyond quantitative requirements such as the number of users to be supported are consideration of the effects that the operating system, commercial off-the-shelf, applications, and communications software have on the hardware. Software issues will be covered in Section 4.1.2.

The capabilities of the workstations used to access the search system must form a seamless interface with the search system. Workstation hardware and software are covered in Sections 4.1.1.5 and 4.1.2.5 respectively.

4.1.1.3 Image System

The image system stores compressed images of all documents in the LSS on optical disks in optical jukeboxes for on-line retrieval and display on Level 2 workstations or for off-line volume printing of documents <u>via</u> high speed laser printers. It is connected directly to the search system from which it receives commands. Output is routed directly to the workstations <u>via</u> the communications system.

The image system consists of three components that are interconnected by a local area network. The components are: (1) image preprocessor; (2) jukebox controller and jukebox storage unit(s); and (3) printer controller and high speed laser printer. Each of these is covered separately below.

The image preprocessor accepts compressed images and associated retrieval data (document accession number, date of issue, and number of pages) from the capture systems. The data can be supplied on magnetic tape, optical disks, or <u>via</u> the communications system. The image preprocessor then performs any translations necessary and presents the image and associated retrieval data to the optical jukebox controller for permanent storage on one of the optical disks in the jukebox storage unit(s).

The jukebox controller and its storage unit(s) are the heart of the image system. The controller is a powerful minicomputer with many functions including:

1) Acceptance of new images from the the image preprocessor. This is a complex event requiring a check for duplication, calculation of the optimal optical platter and the location of that platter in the jukebox storage unit, transfer of the optical platter to a optical disk drive, writing the image to disk and a transaction log, and update of the controller's internal data base of documents and their storage location.

2) Servicing requests for image retrieval and transfer. Based on commands from the search system, the jukebox controller must check its internal data base to verify that it has the document, send a message to the search system if it does not, retrieve the correct optical disk and place it in an optical drive (unless the disk happens to already be in an optical drive), and then transfer the image to the printer controller or to a Level 2 workstation via the communication system. At the workstation the image is decompressed and displayed.

3) Maintaining statistics on usage of specific documents - not by which user, but by the number of times an optical disk or document is accessed. The controller will move optical disks within the jukebox(es) and will group documents often accessed in order to minimize retrieval times.

The printer controller and high speed laser printers print whole documents or pages of documents, plus mailing information, based on commands from the search system. After receiving the print command and associated data, the printer requests the document images from the jukebox controller, temporarily stores them at the printer controller, and then decompresses the images and prints them. The temporary storage is necessary to minimize the service time by the jukebox system and to allow special services such as the printing of multiple copies or on-the-fly reordering of printer priorities. For example, a long print request could be suspended while printing, a high priority short document could be printed and the long print request resumed.

4.1.1.4 Communications System

The LSS communication system must provide ASCII and image data transport at a speed consistent with the response time requirements stated in the Preliminary Needs Analysis. These needs are translated into medium (2.4 to 9.6 Kbps) and high speed ($\geq$ 56 Kbps) data transport between the Level 1 and Level 2 workstations and the co-located search and image systems. The data transport speeds will be established based on the results of modeling efforts, which use parameters such as number of users and the

Level 1/Level 2 workstation mix at each site, access needs as a function of time and program schedule, and request service times by the search and image systems.

The key concept in the communications network topology is one of a unified computing resource. The diverse geographic locations can be viewed as different floors of the same large building. Every LSS user, regardless of location, will view the system as though directly connected to the search system. In the Base Conceptual Design, White Flint is taken to be the communication hub only for illustration. The hub will be co-located with the search and image systems. The concept is achievable through the use of five technologies: local area networks, intelligent bridges, high speed multiplexers, high speed modems for voice grade switched circuits, and intelligent communications processors. Figure 3 shows the network topology with Washington, DC as the location of the search and image systems. High concentrations of users in Washington, DC, Nevada and Texas (see Preliminary Needs Analysis) are supported by LANS while other users are supported through dial up services. Each of these technologies, and their role in the communications system of the Base Conceptual Design, are covered below.

Local area networks, specifically those based on Ethernet, are placed at the concentrated usage locations. Ethernet provides a 10 million bit per second pathway between all the devices attached. In this case, the devices are Level 1 and Level 2 workstations, and, optionally, printers. Software on the Level 1 and Level 2 workstations will allow transparent access to the search system if directly connected (co-located) or connected <u>via</u> a bridge.

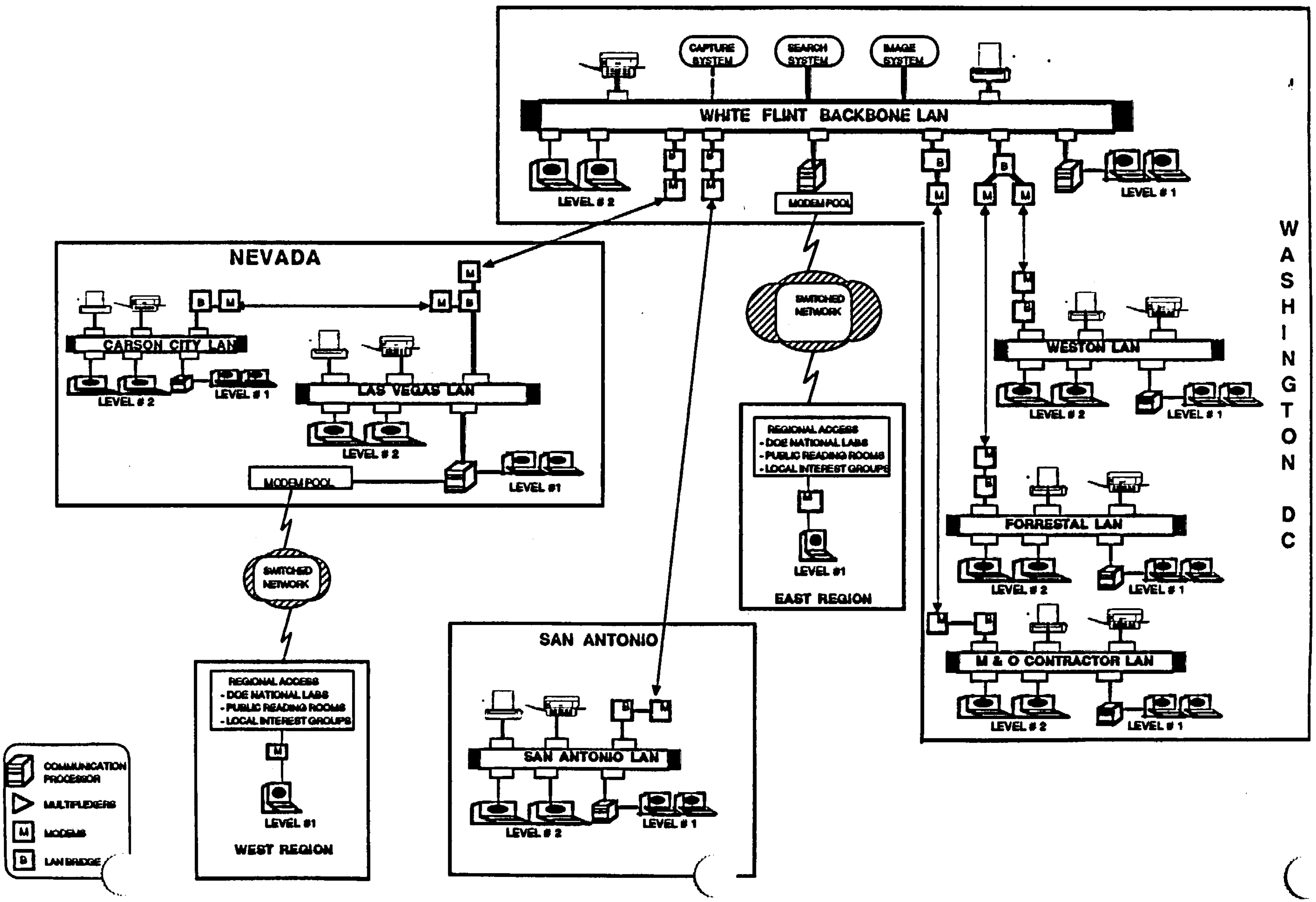
Bridges are intelligent devices that connect two individual local area networks <u>via</u> a wide area network allowing the "bridged" local area networks to appear as a single network. The bridge intelligence is in its ability to route only data from one local area network to another if the data is destined for a particular workstation on the other side of the bridge. This capability becomes important as more local area networks are bridged together. The image system of this conceptual design has its own local area network for communication between the image preprocessor, jukebox controller, printer controller, and the mainframe. A second local area network is supported by the image system for direct and bridged connection to the site local area networks.

High speed multiplexers allow multiple workstations to share a leased communication circuit, each viewing the circuit as though it were dedicated solely to that workstation. High speed modems use new integrated circuit and error correction technologies to send across switched voice grade circuits at speeds such as 9.6 Kbps.

The intelligent communication processors can apply to all the aforementioned technologies simultaneously, providing tailored solutions for each location. In addition, the communication processors monitor the network performance.

FIGURE 3

LSS TOPOLOGICAL COMMUNICATION NETWORK ARCHITECTURE



4.1.1.5 Workstations

The workstations are connected via the communications system to the search system. As described in Section 2.4.1.1, two types of workstations will be supported by the search system. The major difference in the workstations is that Level 1 workstations do not support on-line display of document images, and can only display partial full text ASCII pages. Both would have the same query capabilities and allow LSS users to order the document images for subsequent off-line printing.

The Level 1 workstations are micro computers provided by the users. To be a Level 1 workstations the user device must have the following capability:

1) Dedicated intelligence equivalent to an IBM PC or compatible.

2) Full screen terminal emulation is performed to support the search system software.

3) Downloading of ASCII text from the search system is fully supported if the user has compatible file transfer software.

4) The workstation can also be used for any standard personal computer software applications.

5) A modem is required, which will access the search system via the switched telephone network.

6) Text printers can be attached to a Level 1 workstation to print copies of the screen, reproduce all search system responses and user inputs (record a session in hardcopy), and print text downloaded from the search system during a user session.

Images of documents are available to the Level 1 workstation users only through the print capability of the image system. After retrieving and inspecting a document in ASCII, the user can issue an order for hardcopy, which will be sent by express delivery.

The Level 2 workstation, configured specifically for LSS, has all of the capabilities of the Level 1 workstation, plus the ability to transfer, display, and locally print images. Significant features of the Level 2 workstation are:

1) The Level 2 workstation would be based on a high speed microprocessor that supports multiple parallel tasks, such as the INTEL 80386.

2) Dedicated hardware, including a large format high resolution monitor and image decompression boards and additional software, are used for image display. The large format monitor would also display full page ASCII text.

3) A high speed connection to the communications system supports the transfer of images.

4) A laser printer is available for the printing of images and ASCII text.

Estimates of the number of workstations that would be required at operational startup and at the first peak usage period are shown below.

Number of Workstations

Workstation Level	August, 1990 Min	August, 1990 Max	January, 1994 Min	January, 1994 Max
1	50	100	110	230
2	25	50	65	120
Total	75	150	175	350

These estimates were derived from usage data identified in the Preliminary Data Needs Analysis report and assumptions about which users would need on-line image viewing. The expected number of users during a peak period is 350. The upper limit on the total number of workstations is for each user to have a terminal, with one half of the technical, engineering, regulatory and licensing users requiring on-line images, and thus Level 2 workstations. The minimum estimates would provide one workstation for every two users and only a quarter of the technical, engineering, regulatory and licensing sessions would require on-line images.

4.1.2 Base Conceptual Design Software

Each of the major hardware systems described above require software to perform as an integral part of the total LSS. This software is described on the following sections.

4.1.2.1 Capture Software

There are three major capture functions; image scanning and capture, optical character recognition to obtain full ASCII text, and the capture of cataloging data. The major capture software modules are:

1) Image scanning and capture software - produces files containing the bit-mapped images and page order information.

2) OCR preparation software - obtains the files produced by the scanning process, selects an OCR station, and presents it a page at a time.

3) Check and edit of OCR output - obtains the OCR output, checks the spelling and presents those results and OCR conversion statistics to an operator at a workstation. The operator-corrected ASCII text is written to a file.

4) Cataloging software - the keyed cataloging data of the Records and Regulations Access Subsystems are checked against controlled vocabularies and the thesaurus, and are written in a form ready for data base loading. Software tools that automate portions of the cataloging process may be available.

5) Duplicate document checking - searches the cataloged data in the operational data base to see if a document has already been entered.

6) Cataloging quality control - provides consistency and format checks for cataloging data.

7) Controlled vocabulary and thesaurus maintenance - provides for the addition to controlled vocabularies and the thesaurus.

8) Electronic text input - handles the receipt of ASCII files for documents prepared by word processors and for electronic filings received via E-mail.

4.1.2.2 Search Software

The search software, runs on the LSS search hardware and implements the main LSS function - computerized access to its data. It includes:

1) Applications software to implement the LSS search and retrieval needs that have been and will be identified

2) Commercial data base management system software to implement the LSS data base design, and provide many of the query, retrieval, and utility functions

3) Operating system software, provided by the manufacturer of the computer hardware, which controls and schedules all of the jobs running on the computer hardware.

To satisfy the LSS search and retrieval needs, the application software will consist of a number of modules, each related to a specific function. The major modules will be:

1) Menus - One method a user will be able to choose what they want to do, is to select from a list of choices displayed in menus. The choices presented will likely range from a list of the LSS subsystems to specific search criteria. The menu module will also manage entry to and linking of the other modules. Menu navigation will permit knowledgeable users to go directly to any menu, whether or not it appears as choice on the menu currently displayed.

2) Query screens - Fill-in-the-blank type of screen displays which support the formulation of query requests for the Records and Regulation Access headers, and the information in the Tracking Subsystems. These screens will likely also incorporate windows

which can be used to view items such as controlled vocabularies, the thesaurus, indices and previous result reports. Special query screens may also be developed for accessing the full text.

3) Prompting dialog - This module would "converse" with a user to construct a query to pass to the data base management system. It would provide the ability to easily store and modify an earlier request. The module accesses both the headers and full text associated with the Records and Regulations Access Subsystems.

4) On-line help - Interfaces with the other modules to answer user questions and to display informative and helpful error messages.

5) Image display and hardcopy requests - Obtains a request from Level 2 Workstation user for the display or printing at the workstation of one or more images, passes the request to the Image System Link module. Marking of specific pages viewed <u>via</u> full text may be one method of selecting pages to be printed.

6) Image system link - Passes request obtained by the image request module to the image system along with the information the communication system needs to route the image to the proper workstation. If the image system also processes the requests for hardcopies that are to be mailed, then this module will also pass those requests to the image system.

7) SIMS interface - Provides the software and data base link between the Records Access Subsystem and the Sample Inventory Management System (SIMS).

8) Queries and results save and reuse - Will be used by the prompting dialog and query screens modules to enable users to save queries they have constructed for future reuse or as a basis for preparing a new query. It will also save and identify result sets for later display or a restructure for a future query.

9) Parallel query processor - The base design assumes that the full text data base can best be implemented by partitioning it into several data bases. This module would make the partitioning as invisible as possible to the user. On the query side, it would invoke the same query on each partition. On the retrieval side, it would combine result sets.

10) Data base loading - Controls and schedules the entire data base loading process. The actual loading will be done <u>via</u> the loading utilities provided by the data base management system. Other functions of this module include synchronization with the image loading process and providing information for menu system to display data concerning the contents of the data base.

11) Performance monitoring - Collects statistics for the LSS administrator such as the type and complexity of queries generated <u>via</u> other modules, the size of result sets obtained, the time required for classes of queries to be completed and amount of text

and images displayed. This data will be used to optimize system performance.

12) Configuration management - Will contain data describing the hardware and software configuration of the LSS and will track changes made to the baseline. This module will be implemented using a commercial software package.

The data base management system software is the key component of the LSS search software. Its ability to handle data bases of the size projected for LSS and to support the following types of query and retrieval functions is crucial for the success of the LSS. The following functions need to be supported by the data base management system:

1) Store and invert indices for up to 200 Gbytes of full text (upper estimate for the year 2009)

2) Implement both full text and relational data base structures

3) Support the following query constructs
   - High level, easy to use full screen query tools
   - Full screen forms
   - Boolean expressions
   - Stem wildcard searches
   - Proximity word searches
   - Near spelling or spelling correction

4) Optimization of searches, including segmented refinement of search in order to narrow retrieval sets

5) Thesaurus support for retrieval

6) Simple and complex user report generator(s)

7) On-Line help facility

8) Utilities
   - Fast, large volume loading
   - Backup and recovery
   - Integrity checking
   - Usage monitoring
   - Tuning

4.1.2.3 Image System Software

The image system consists of six components - image preprocessor, jukebox controller, jukebox(es), printer controller, printer(s), and local area network. Three of these (the image preprocessor, jukebox controller, and printer controller) have embedded microprocessors that require software. Each is described below.

The image preprocessor has software that:

1) Accepts input image data on magnetic tapes and optical disks

2) Translates, as necessary, to vendor compressed image format and forward to the jukebox controller.

Software for the jukebox controller:

1) Checks for duplication of an existing document by accession number

2) Calculates the optimal optical platter for storage of the image and its location

3) Writes the image to the chosen optical disk and updates the internal location data base

4) Responds to requests from the search system and printer controller.

The printer controller software functions include:

1) Receiving requests from the search system for printed images

2) Requesting and receiving images from the jukebox controller

3) Buffering images to magnetic disks as necessary

4) Image decompression and transfer to high speed laser image printers.

Depending on the system chosen, the vendor supplied software could range from primitive subsets to completely integrated packages with high level control languages.

4.1.2.4 Communications Software

Each of the systems in the Base Conceptual Design must communicate with at least one of the other systems. A combination of vendor supplied and custom designed software will be required for integration of the capture, search, image, communication, and workstation systems. The seven layer Open System Interconnection (OSI) standard will be followed wherever possible. The level of custom software required will depend on the final vendors chosen for the systems. Regardless of the selected vendors, there are key functions for each system that must be performed. Each system is addressed separately below.

The capture system minicomputer must be able to communicate with the optical character recognition (OCR) machine, the optical disk drive, the scanners, and the search system. The OCR is sent images and returns ASCII text. The optical disk drive requires the transmission of commands and images from the minicomputer. Software that accepts data from the scanners and writes the images to local storage is required. Communication with the search system will be *via* magnetic tape, optical disks, or direct connection. In the case of a direct connection, the capture system minicomputer will have software which performs any protocol conversion for transmission of text and images to the search system.

The search system will accept data from the capture system, transmit commands to the image system, and interface with the workstations. Separate software components will be required for each interface. All of this communication will be simultaneous yet transparent to the LSS user.

The software to support the electronic mail functions described in Section 2.6 would also operate on the search system hardware and be interfaced with the data transfer and data capture software. The integration would enable messages in filings to be composed on a local workstation editor and then sent to the E-mail software for distribution. When appropriate, these items would be forwarded to the data capture software.

The image system components will communicate with each other. The jukebox controller and printer controller will accept commands from the search system, return status information to the search system, and transmit images to the Level 2 workstations.

The workstations will communicate with the search system by emulating terminals and through downloading from the host. Level 2 workstations will also communicate with the image system, accepting compressed images and decompressing them for display and local printing.

Table 3 summarizes the high level communications software functionality required. Some custom software will be required in many cases to interface the vendor supplied software to the LSS applications software. This effort may be complex and, depend on the functionality of vendor equipment and software, especially when multiple vendors are used.

4.1.2.5 Workstation Software

The two levels of workstations for query users each will have some local software operating that enables them to interact with the rest of the LSS retrieval software and hardware.

The Level 1 workstation will require the following software, both of which are commercial products:

1) Terminal emulator - allows the workstation to act as a terminal to the search computer. The type of terminal to be emulated depends on what terminals the search computer and software will support.

2) Data transfer - for receiving data from the search computer. This needs to be compatible with the LSS communication software. This function is sometimes included in terminal emulation packages. For a Level 1 workstation, this function is optional.

The Level 2 workstation software includes the Level 1 software plus the following:

1) Image handler - receives the images from the communications system, controls the display of a image on the workstation's full-page display, and routes images to the local laser printer.

TABLE 3. SOFTWARE REQUIRED FOR LSS COMMUNICATIONS

System	Function	Vendor Supplied Software	Custom Developed Software
Capture:	Communicate between OCR and minicomputer	X	
	Communicate with optical disk drive(s)	X	D
	Communicate with scanners	D	
	Communicate with search system	X	D
Search:	Communicate with capture system	X	D
	Communicate with image system	D	X
	Communicate with workstations	X	D
Image:	Communicate with search system	D	X
	Communicate with workstations	D	D
Workstations:	Communicate with search system	X	
	Communicate with image system	D	D

X - Will be required regardless of vendor supplied software.
D - Depends on the hardware vendors chosen.

2) Multi-tasking operating system - such operating system software will-allow the image printing to occur in parallel with continuing query activities being conducted _via_ the terminal emulation software.

4.2 Variants to Base Conceptual Design

Several major features of the LSS may be implemented in more than one viable way. This section presents seven variants on the Base Conceptual Design that provide such alternatives. Although all variants satisfy the functional requirements, the efficiency and manner in which they are satisfied will vary.

4.2.1 Variant I - Full Replicated Nodes

4.2.1.1 Description

Variant I (Figure 4) differs from the Base Conceptual Design in that it has two fully replicated search and image system nodes, one located in Washington, DC and one in Las Vegas, NV. The system architecture consists of the identical hardware and software configuration as the Base Conceptual Design. This variant was chosen to exploit the data security that full redundancy offers from a backup and restore perspective. Additional advantages include higher overall availability if the centers are linked plus potentially lower recurring communication charges, since the leased lines required would be shorter in total. Partial replication or distribution of the data base were also considered but not selected as variants because the required complexity of the associated query software and the severe difficulties of data configuration management.

4.2.1.2 Impact On The Capture System

There are no changes in the hardware and/or software configuration of the capture system. The text and images created will be duplicated for input and storage at the two nodes. This additional step will somewhat lengthen the total processing time.

4.2.1.3 Impact On The Search System

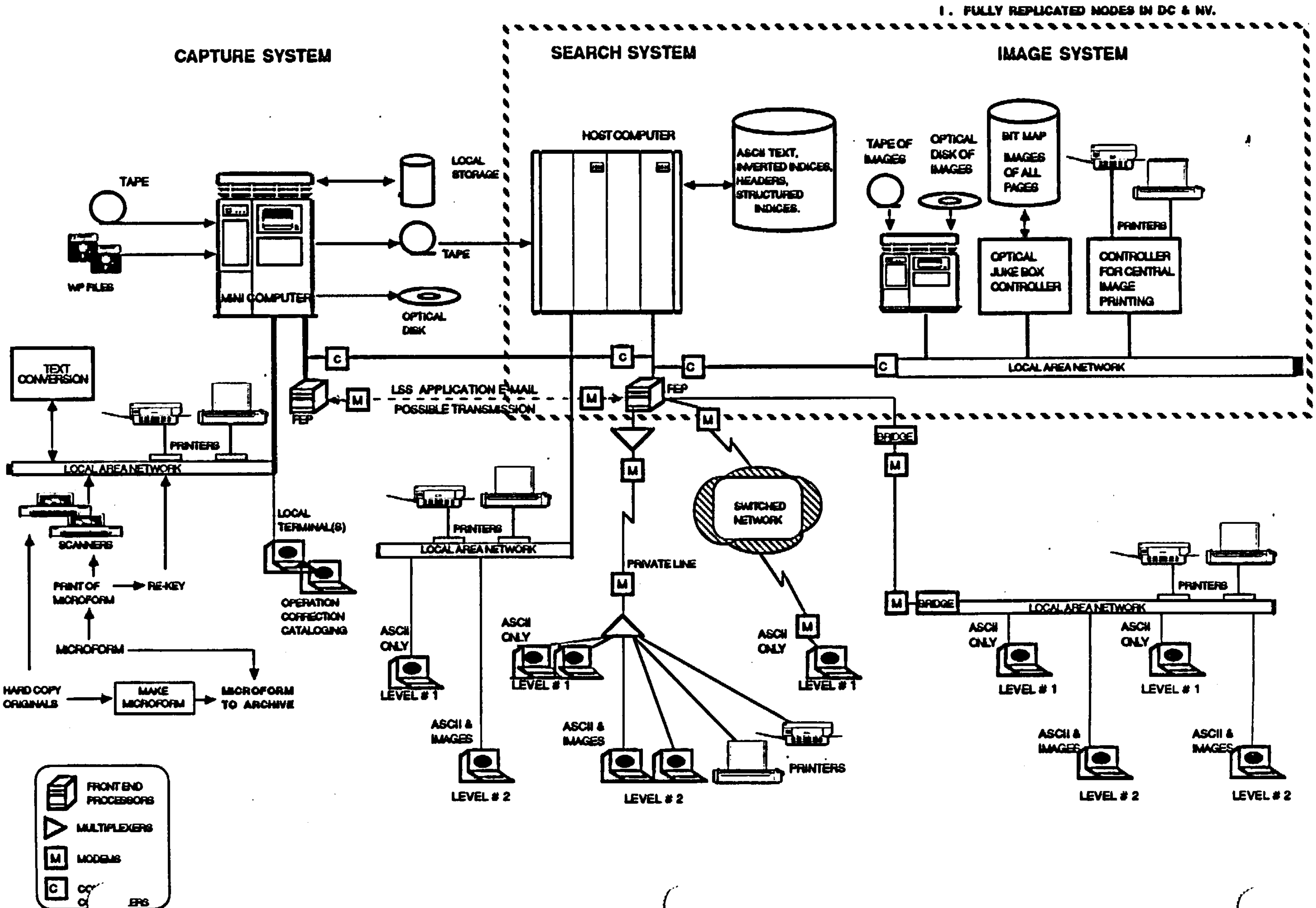
The hardware and software configuration for the two-node system will be identical to the one-node Base Conceptual Design architecture. The LSS data bases are replicated and stored at both nodes. There is an increased staffing and configuration management overhead in keeping the two centers synchronized.

4.2.1.4 Impact On The Image System

There will be identical image systems co-located with the search systems at both Washington, DC and Las Vegas. All LSS images are stored

FIGURE 4

VARIANT I



at both centers. The same configuration management problems affecting the search system also apply here.

4.2.1.5 Impact On The Communications System

The network will be geographically split between the Las Vegas and Washington nodes. The communications traffic associated with Las Vegas region (including Carson City) will be supported by Las Vegas. All of the Washington area users will be supported by Washington. San Antonio could be supported by either node. The communications processors at each node would be interconnected to allow transmission of E-mail and failover support if one center should go off-line.

4.2.1.6 Impact On The Workstations

Level 1 and Level 2 workstation configurations are unaltered from the Base Conceptual Design.

4.2.2 Variant II - Hardware Full Text Search

4.2.2.1 Description

As previously described in Section 4.1.2.2, in the Base Conceptual Design full-text search is implemented <u>via</u> storage of the full text plus creation of an inverted index. The full-text data base software uses both to respond to user queries. Variant II (Figure 5) replaces them and the text data base manager with specialized hardware processors. The hardware compresses the full text and stores it on very high transfer rate disk drives for subsequent searching. The search is performed serially through <u>all</u> of the compressed full text. The search speed is a function of data base size, disk transfer speed, and the relative size of the query relative to the width of a comparator. The hardware full-text processor is directly attached to the search host computer from which it receives search requests and returns decompressed full text. Multiple hardware processors can be connected to the host computer. This allows the system to maintain a consistent search time as the data base grows. All other functions (E-mail, header searches, <u>etc</u>.) performed by the search system host computer will remain the same as the Base Conceptual Design.

This variant was chosen in order to capitalize on four advantages offered by hardware full-text search:

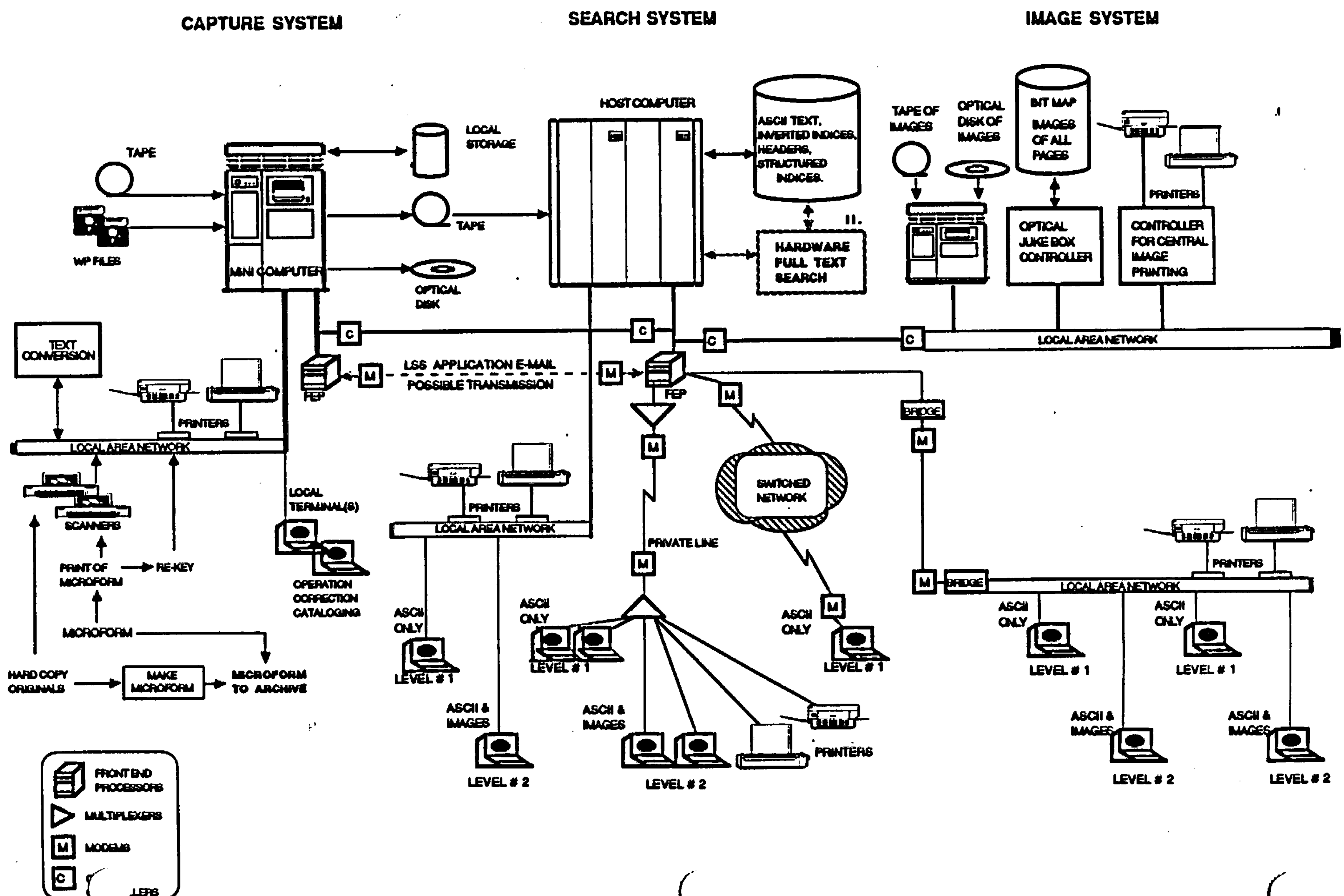
1) A 4:1 reduction in the amount of required disk storage. This is a result of the average 50% compression achieved by the hardware full-text processor <u>vs.</u> at least a 100% expansion (full text plus inverted index) for the software full-text solution.

2) Predictable full-text search response times.

3) Update simplicity since new documents are simply added to the end of the data base and no index creation or update is required.

FIGURE 5

VARIANT II



4) Over the long operational life of the LSS hardware full-text search technology is expected to advance much faster than software technology.

4.2.2.2 Impact On The Capture System

There are no impacts on the capture system, compared to the Base Conceptual Design.

4.2.2.3 Impact On The Search System

Currently available vendor supplied software for the hardware processors is primitive compared with the current state of the art in software data base managers. The result is a significant increase in the software effort and complexity, compared to the Base Conceptual Design, particularly in the software required to coordinate multiple hardware processors.

4.2.2.4 Other Impacts

There are no impacts in Variant II compared to the Base Conceptual Design on the image system, communications and workstations.

4.2.3 Variant III - Images Are Not Supported At Workstations

4.2.3.1 Description

In this variant (Figure 6) from the Base Conceptual Design, the capability to view electronic (bit-mapped) images on the screen at the Level 2 workstations is excluded. This variant was selected since this capability was not identified as necessary to all potential LSS users in the Preliminary Needs Analysis.

4.2.3.2 Impact On The Image System

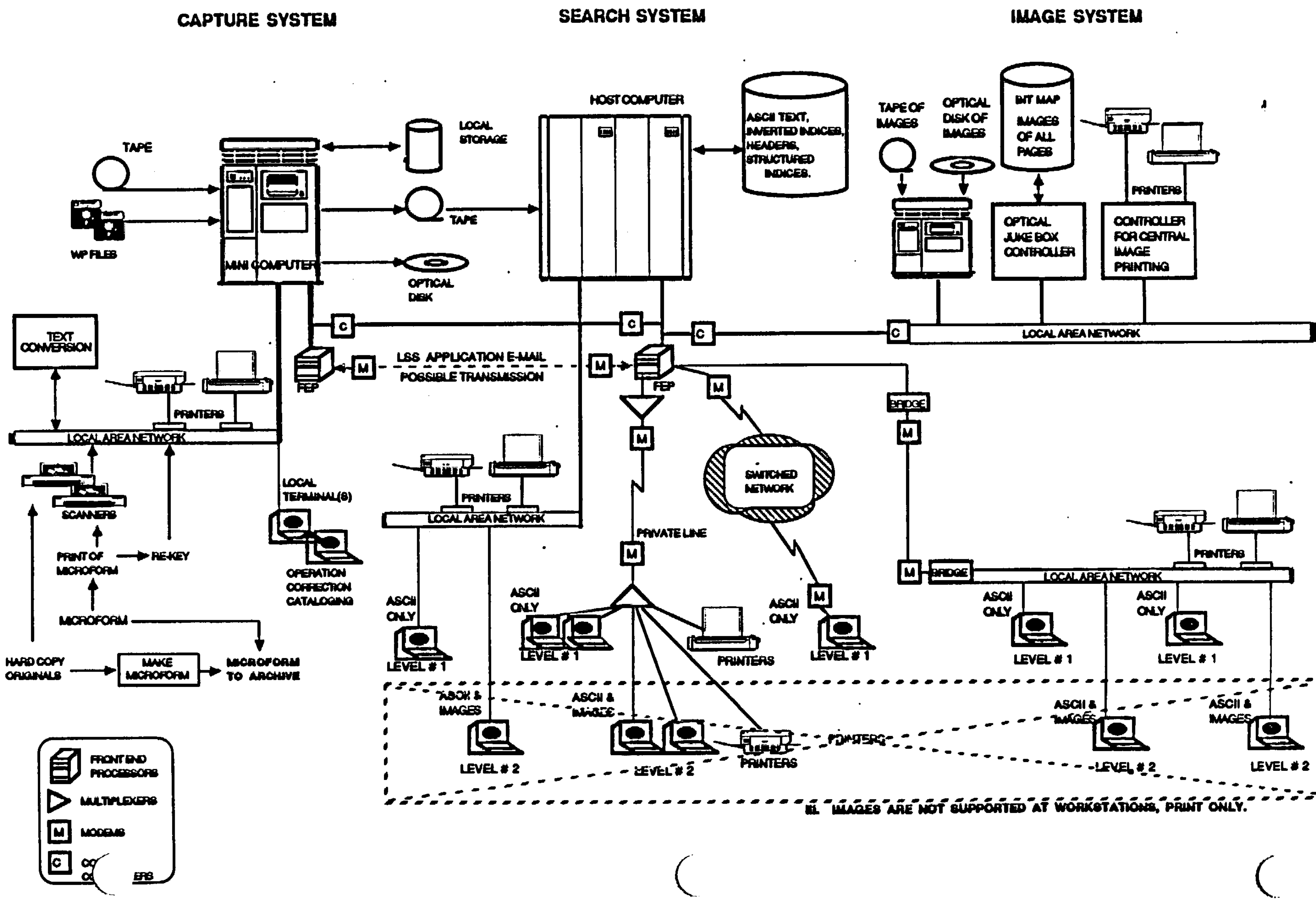
The software to transmit images stored in the image system to Level 2 workstations would not be required by this variant. There is a decrease in load on this system resulting from this reduced capability. To compensate for not being able to display images at a workstation, the requests for printing images would increase.

4.2.3.3 Impact On The Communications System

The elimination of the transmission of bit-mapped images to Level 2 workstations reduces the need for high capacity telecommunications links to user locations.

FIGURE 6

VARIANT III



4.2.3.4 Impact On The Workstations

Variant III eliminates the ability to view images on-line. This will result in the elimination of image hardware and software for Level 2 workstations. Level 2 workstations would still be able to display full page ASCII text.

4.2.3.5 Other Impacts

There are no impacts in Variant III compared to the Base Conceptual Design on the capture and search systems.

4.2.4 Variant IV - Microform Digitizers in Capture and Image Systems

4.2.4.1 Description

In this variant (Figure 7) on the Base Conceptual Design, the changes occur in the capture and image systems. Microform digitizers are used to create the OCR input for documents available only on microform. Microform replaces optical disks for the storage and retrieval of images. This variant was chosen to reflect the availability of automated microform systems.

4.2.4.2 Impact On The Capture System

A microform digitizer is added to the capture system to convert documents available only as a microform image to a bit-map similar to the bit maps of hardcopy documents generated by scanners. From this point, conversion to text using the OCR device is the same as in the Base Conceptual Design. Microform images not judged acceptable because of poor film quality will be printed to hardcopy and text re-keyed from the hardcopy. After scanning or re-key, the microform is sent to the microform image retrieval system described in Section 4.2.4.4. Documents that arrive in hardcopy are processed by the capture system as in the Base Conceptual Design.

4.2.4.3 Impact On The Search System

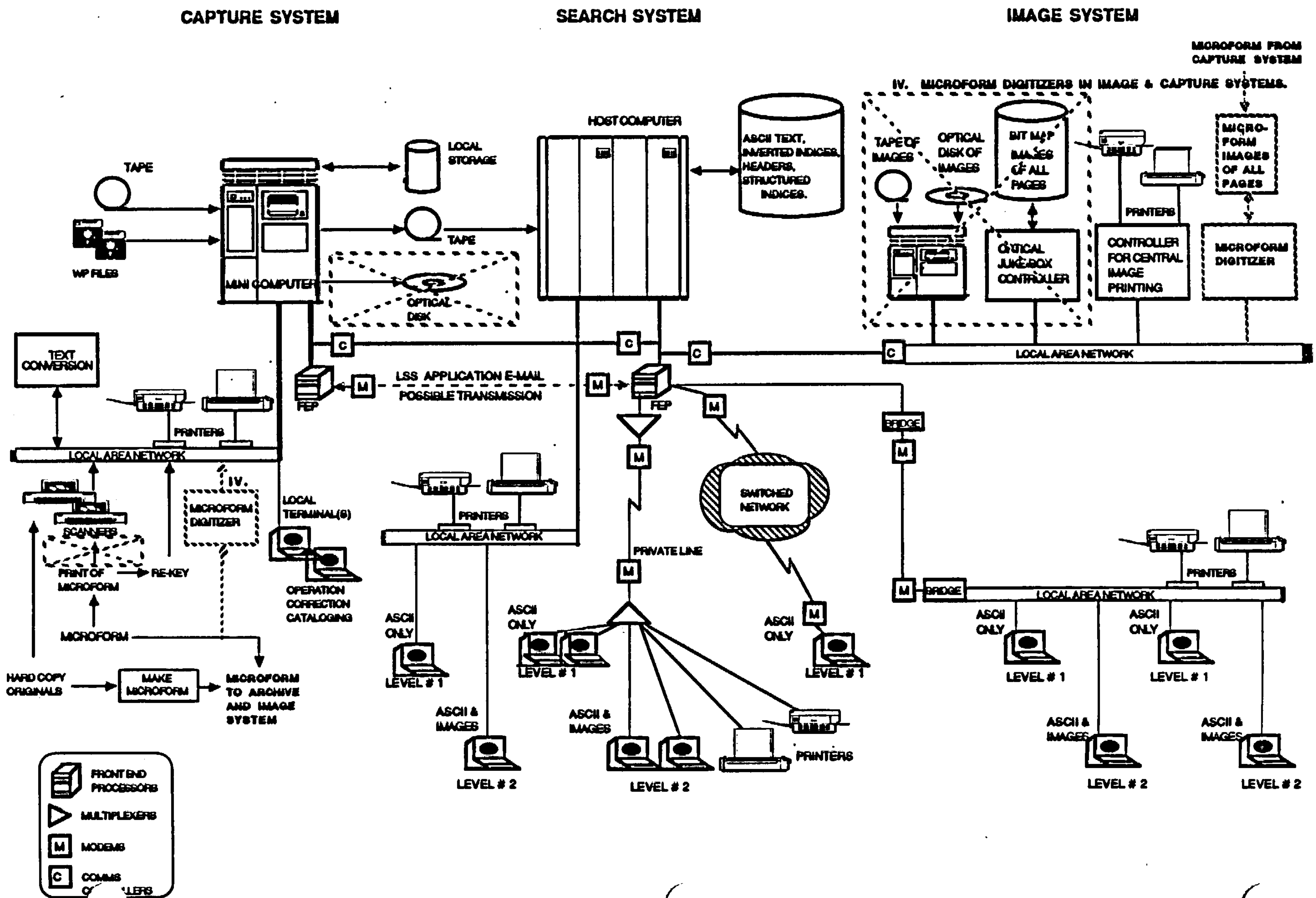
The image system link of the search system software must interface with the microform-based image retrieval system in this variant, as opposed to an optical disk system in the Base Conceptual Design. No other impact on the search system results.

4.2.4.4 Impact On The Image System

The optical preprocessor, jukebox controller, and optical jukebox components are replaced in this variant by a microform on-line storage and retrieval system which retrieves, digitizes, compresses, and transmits the microform image to a Level 2 workstation or image printing controller. The

FIGURE 7

VARIANT IV



software required to control the automated microform retrieval and digitizing system will be different from the optical jukebox but does not significantly differ in complexity.

As opposed to the optical disk base conceptual design, in which the storage medium requires no maintenance, a microform-based image storage system can require the replacement of all the storage medium in the system on a frequent basis (depending of specific usage levels).

4.2.4.5 Impact On The Communications System

There should be no impact on the communications system. Compressed images will be of approximately the same size and, depending on the number of microform retrieval units used, will be transmitted at the same frequency.

4.2.4.6 Impact On The Workstations

The Level 1 users will not be affected. Microform stored in such a system deteriorates with age and handling. The Level 2 user may notice minor variations in the quality of the same image if it is retrieved multiple times since the microform image is digitized each time it is retrieved.

4.2.5 Variant V - Microform Off-Line Image Storage and Retrieval

4.2.5.1 Description

Variant V (Figure 8) of the Base Conceptual Design replaces the on-line image system with a off-line service for obtaining hardcopy or microform copies of LSS documents. This is similar to the way commercial and existing DOE bibliographic data base services provide document copies to their users. For example, DIALOG allows users to order documents from NTIS as a command after locating the document.

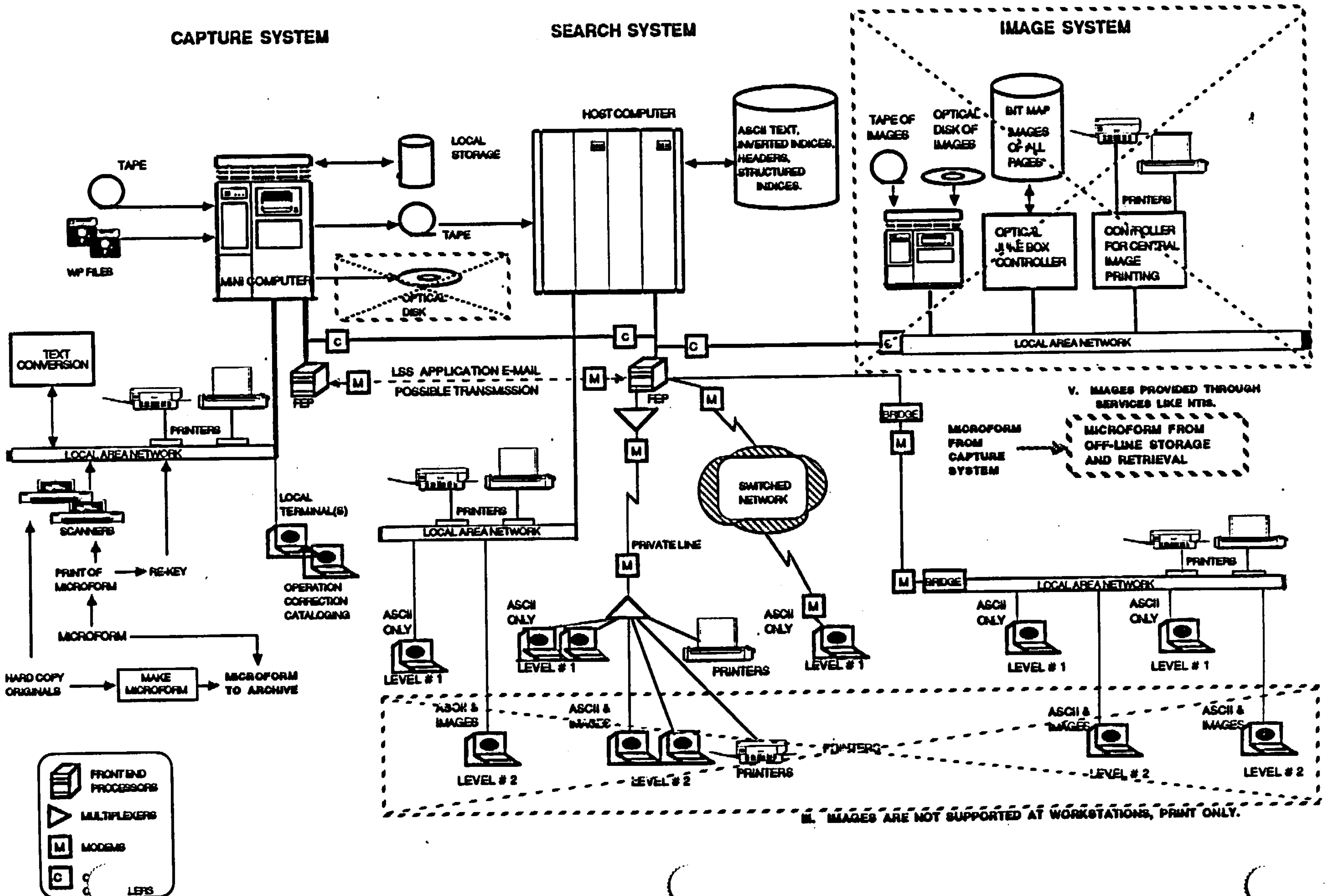
This variant was developed to present a low-tech solution to meeting the hardcopy receipt time requirements of 2 to 3 days identified in the Preliminary Needs Analysis. The function capability to view electronic (bit-mapped) images on the screen at the Level 2 workstations is excluded, as in Variant III.

4.2.5.2 Impact On The Capture System

The capture system no longer captures the images to optical disks. A duplicate of the microform is sent to the off-line service, which operates similar to the NTIS, described below.

The National Technical Information Service (NTIS) maintains government and contractor documents on microfiche. In addition, some documents are available in hardcopy and on computer tape or diskette. Copies of documents

FIGURE 8
VARIANT V



can be requested by telephone, by mail, or electronically. Documents can be requested as a rush order, in which case the document will be ready for pickup or mailing within 24 hours.

Upon receiving a request the corresponding microfiche is manually retrieved from the archives. Depending upon the request, the microfiche is either duplicated or blown back into paper form. The original is returned to the archive, and the copy is packaged and prepared for delivery. The packaged copy can be either picked up in person at several locations in the Washington, DC area during office hours or shipped to the requester through the postal service or through an express carrier if requested.

4.2.5.3 Impact On the Search System

The search system no longer has to interface with and control the image system. It simply takes orders for documents and forwards them to the off-line service. The elimination of the control software for the image system is significant in terms of personnel and schedule risk reductions.

4.2.5.4 Impact On The Image System

The image system from the Base Conceptual Design is completely eliminated, replaced by the microform off-line system described above.

4.2.5.5 Impact On The Communications System

The elimination of the transmission of images to Level 2 workstations reduces the need for high capacity telecommunication links to the user locations.

4.2.5.6 Impact On The Workstations

The elimination of the on-line image system results in a similar hardware and software change for the Level 2 workstation, as in Variant III.

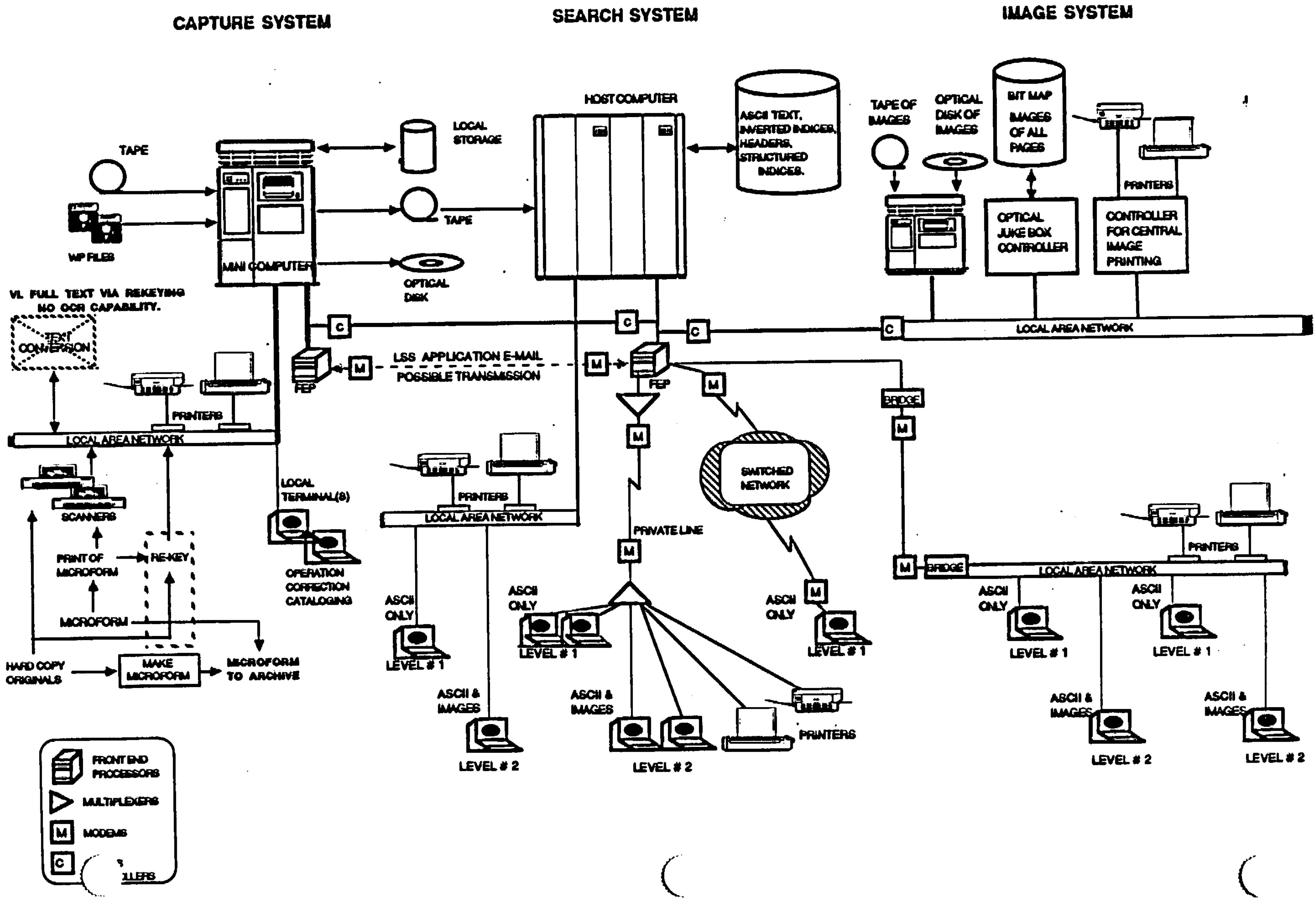
4.2.6 Variant VI - Full Text *via* Re-keying

4.2.6.1 Description

In this variant (Figure 9), there is no automated text conversion (OCR) process. The conversion of hardcopy text to ASCII is accomplished by re-keying the document. An expected 99% accuracy of data *via* re-keying would be achieved by double keying the original source document. There is a new requirement for software to assist in management and control of the re-keyed documents and bit-mapped image documents. The software should keep track of the documents, location to which they are sent, which have been re-keyed and returned, which have passed the QA process, the status of the image processing and the status of cataloging, so that all of a documents' components are available for loading into the search and image systems.

FIGURE 9

VARIANT VI



This software will increase the integrity of the process, and reduce the possibility of document duplication.

4.2.6.2 Impact On The Capture System

Since the text conversion will be accomplished via re-keying there will be no requirement for optical character recognition equipment, and associated software. However, the re-keyed documents would require processing through a digital scanning device since bit-mapped image capture and storage is required. (The OCR process allows for the simultaneous conversion of text and digitization of images). There will be a requirement for additional software to assist in the configuration management.

4.2.6.3 Other Impacts

The search system, image system, communication, and workstations in Variant VI are the same as in the Base Conceptual Design.

4.2.7 Variant VII - Combined Variants III, V and VI

4.2.7.1 Description

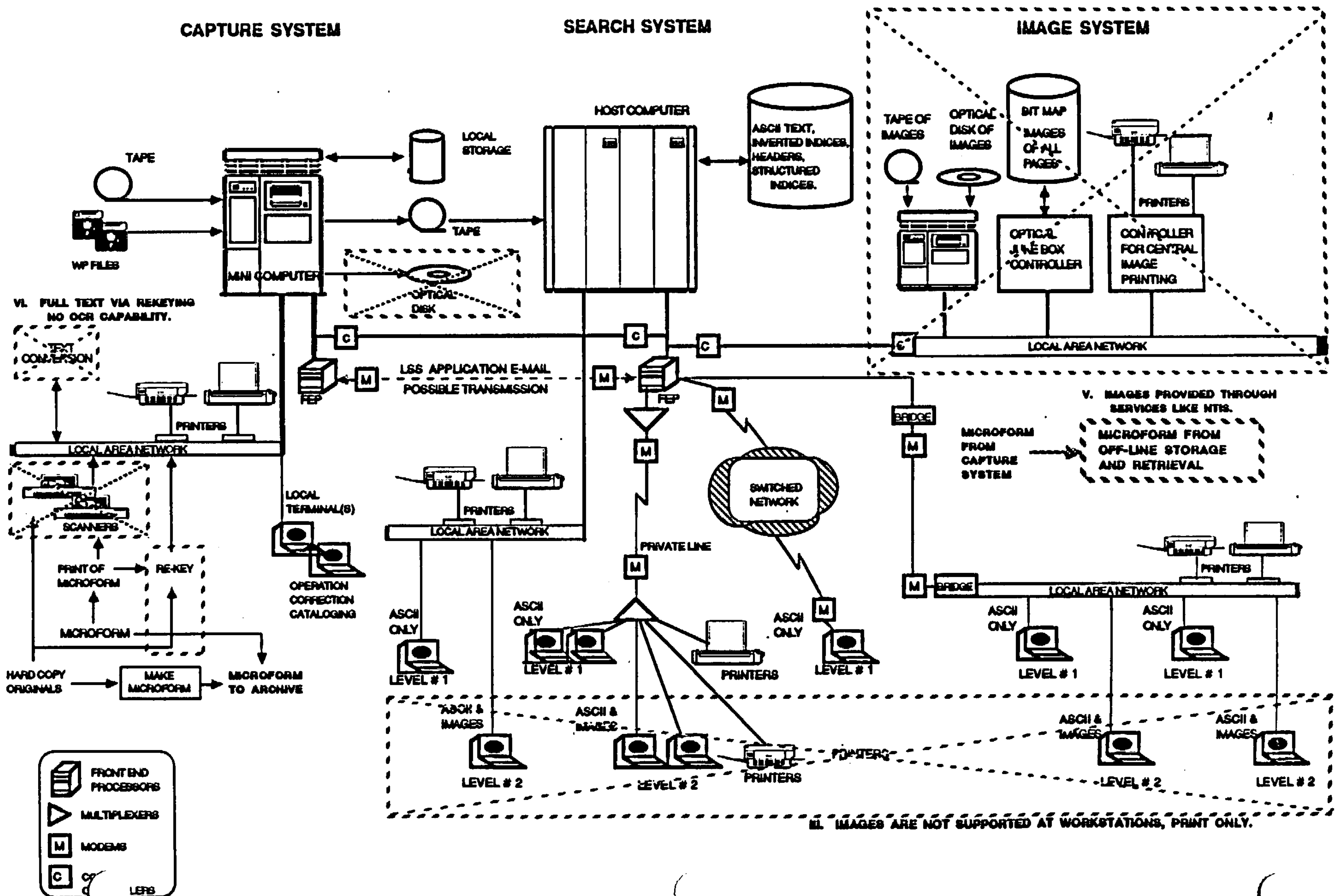
Variant VII (Figure 10) combines hardware and software changes for the Level 2 workstations (Variant III), the removal of the on-line image system (Variant V) and the re-keying of all documents instead of OCR (Variant VI). This variant was created to present a conceptual design with the lowest schedule risk that minimally meets the requirements presented in the Preliminary Needs Analysis.

4.2.7.2 Other Impacts

The impact on the capture system is the same as Variant VI. The impacts on the search system, image system and communications are the same as Variant V. The impact on workstations is the same as Variant III.

FIGURE 10

VARIANT VII



5.0 CONCLUSIONS

The Preliminary Needs Analysis and Preliminary Data Scope Analysis have defined the requirements of an automated computer-based information storage and retrieval system that must accommodate millions of documents. Based on these studies and the directions perceived from the LSS negotiated rulemaking process, a conceptual design has been formulated which meets the stated requirements. While the design was selected because it represents a comprehensive low risk technical solution, it also represents both a large and complex system. As noted in Section 4.0, this design is the formulation of a number of technical experts and is based on the combined experience of the group. In this process a number of alternatives were examined. Some were rejected due to a low probability of success. Others not only offered a reasonable risk but also were potentially more cost effective, although they may not meet all of the "non-firm" requirements identified in the Preliminary Needs Analysis. These alternatives were identified as variants to the base and will be subject to further technical and economic investigation.

The Base Conceptual Design and variants are consistent with the requirements identified to date, including the deliberations of the Negotiated Rulemaking Advisory Committee. Indeed, the rulemaking activities have not yet imposed any requirements on the design which were not anticipated in the Preliminary Needs Analysis. It is further not expected that the rulemaking activities will result in any requirements which cannot be met by the base design or one of the variants; however, the possibility still exists that design refinements may be required to reflect changing requirements. One probable design refinement is for the Tracking Systems needed by DOE to be implemented by DOE only.

Of perhaps more significance to the refinement of the design is the feed-back from potential users and the cost-benefit analysis. All of the dialogue conducted to this point with potential users of the LSS has been in the absence of a common reference design. This has caused much difficulty in communication, because while one must speak from a certain frame of reference, it is almost assuredly not the same as that of the other party. With this study, a reference point is established for the LSS conceptual design that will facilitate future deliberations. Refinement of the design can benefit from additional experiences of and feed-back from potential users, especially in the area of features available to the user. With this base frame of reference, the potential user can now envision certain scenarios of LSS applications and can fill in the various details necessary to more accurately predict system sizing and response.

The Benefit-Cost Analysis, the next report in the series, will permit the investigation of the variants from the standpoint of cost (both savings and additions), and the associated benefits (or lack thereof) of each variant. Such information will provide an economic basis for decisions on how best to meet the various requirements of the system.

The opportunity is now available for substantive input to the design process, which is both encouraged and necessary to refine the process further.

REFERENCES

ANSI/ASME NQA-1, 1983; Quality Assurance Program Requirements for Nuclear Facilities, American National Standards Institute and American Society of Mechanical Engineers, 1 July 1983.

DOE, 1986; Quality Assurance Plan for High-Level Radioactive Waste Repositories, DOE/RW-0095, DOE OCRWM Office of Geologic Repositories, August 1986.

DOE, 1987; Request for Proposal: Design and Implementation of a Licensing Support System, DOE Office of Civilian Radioactive Waste Management, 11 February 1987.

DOE, 1988a; Licensing Support System Preliminary Needs Analysis, OCRWM, Office of Resource Management, February 1988.

DOE, 1988b; Licensing Support System Preliminary Data Scope Analysis, OCRWM, Office of Resource Management, April, 1988.

APPENDIX A

EXAMPLE OF LSS USER SESSION SCENARIO

Technical & Engineering Initial Background Research Scenario

Session Duration:	2 to 3 hours
Session Frequency:	3 to 5 sessions per study 1 to 2 studies per year per scientist/engineer
Data Accessed:	Technical reports in the Records Access Subsystem

In this scenario the scientist or engineer is using an LSS workstation near (or on the same floor as) their office to perform background research on a technical subject with which they wish to become familiar. The user begins with a list of terms and subjects relevant to the topic and possibly a list of important authors, as well as some idea of the time frame for research of interest. The scientist or engineer begins by working primarily with catalog data. Specific classes of documents, e.g., memoranda, letters, progress reports, can be excluded from the search. Publication or issue date can be used to define a period of interest.

Using primarily subject and/or keywords, terms in the subject field, the user will perform an initial search, view only the number of documents selected (but not yet the catalog information itself), and sequentially refine the retrieval set until the number of selected documents is in the range of 50 to 100. During this process, the user may refer to the thesaurus to find legal keywords.

Initial queries might contain logical ORs, while the result set refinement would be more likely to contain ANDs and NOTs. From time to time the user would be likely to request intersection (ANDing) of the results of two independent sequential refinement searches. Next, histograms by subject and date may be requested.

When the result set is reduced to a set of 50 to 100 documents, the user may change strategies and begin to organize information by author name, and may request to see a histogram of the count of hits by author or may restrict the search to include or exclude selected authors. Only now will catalog information (probably sorted by author) be displayed, both as descriptive and subject catalog information will be requested, followed by abstract text for selected documents.

Having reviewed several documents, the scientist/engineer may change strategies and perform a series of full-text searches using proximity searches with 3 to 5 terms per query. These searches will be performed against a previously-selected subset of documents.

The scientists/engineer will review the text of selected documents at the workstation, typically looking first at the title page, possibly the table of contents, then skipping to the end of the document to review conclusions, references, and appendices. The reader will page both forward and backward through the document and view equations, figures and diagrams as well as text. From 10 to 20 documents per session may be reviewed in such a manner.

The user may request up to 50 pages of printed material perhaps once during the session. Overnight delivery is acceptable. The user will want to save intermediate results from one session to another.

APPENDIX B

REVISED PROJECTION OF THE SIZE OF THE LSS DATA BASE, 1990 - 2009

This appendix contains a revised version of Table 8 (Projected Size of the LSS Data Base, 1990-2009) of the Preliminary Data Scope Analysis. Revisions have been made for the estimates of pages added during 1993 and 1998, based on a re-evaluation of the expected levels of activity consistent with the methods described in that report, and include both a low and a high estimate. The revised values are somewhat higher than those published in the Preliminary Data Scope Analysis.

TABLE 8. PROJECTION OF THE SIZE OF THE LSS DATA BASE, 1990 - 2009

	LOW ESTIMATE		HIGH ESTIMATE	
Year	Pages Added During Year	Cumulative Pages At Year-End	Pages Added During Year	Cumulative Pages At Year-End
1990	830,000	9,304,000	1,100,000	11,885,000
1991	1,087,000	10,391,000	1,441,000	13,326,000
1992	1,428,000	11,819,000	1,892,000	15,218,000
1993	1,660,000	13,479,000	2,200,000	17,418,000
1994	2,009,000	15,488,000	2,662,000	20,080,000
1995	1,858,000	17,346,000	2,463,000	22,543,000
1996	1,635,000	18,981,000	2,167,000	24,710,000
1997	1,386,000	20,367,000	1,837,000	26,547,000
1998	1,037,000	21,404,000	1,374,000	27,921,000
1999	1,286,000	22,690,000	1,704,000	29,625,000
2000	1,170,000	23,860,000	1,550,000	31,175,000
2001	1,877,000	25,737,000	2,487,000	33,662,000
2002	1,236,000	26,973,000	1,638,000	35,300,000
2003	1,261,000	28,234,000	1,671,000	36,971,000
2004	1,327,000	29,561,000	1,759,000	38,730,000
2005	1,120,000	30,681,000	1,484,000	40,214,000
2006	415,000	31,096,000	550,000	40,764,000
2007	365,000	31,461,000	484,000	41,248,000
2008	365,000	31,826,000	484,000	41,732,000
2009	365,000	32,191,000	484,000	42,216,000

APPENDIX C

ABBREVIATIONS USED

ANSI	American National Standards Institute
ASCII	American Standard Code for Information Interchange
ASME	American Society of Mechanical Engineers
bps	bits per second
CD-ROM	Compact Disk Read Only Memory
CFR	Code of Federal Regulations
DOE	Department of Energy
dpi	dot per inch
FFRDC	Federally Funded Research and Development Center
LAN	Local Area Network
LSS	Licensing Support System
M&O	Management and Operations contractor
NNWSI	Nevada Nuclear Waste Storage Investigations
NRAC	Negotiated Rulemaking Advisory Committee, officially known as the HLW Licensing Support System Advisory Committee
NRC	Nuclear Regulatory Commission
NTIS	National Technical Information Service
NWPA	Nuclear Waste Policy Act of 1982
OCR	Optical Character Recognition
OCRWM	DOE Office of Civilian Radioactive Waste Management
OMB	Office of Management and Budget
OSI	Open System Interconnection
PC	Personal Computer

ABBREVIATIONS USED
(continued)

QA	Quality Assurance
QC	Quality Control
RFP	Request For Proposal
SCP	Site Characterization Plan
SIMS	Sample Inventory Management System

LICENSING SUPPORT SYSTEM
BENEFIT-COST ANALYSIS

SCIENCE APPLICATIONS INTERNATIONAL CORPORATION

LICENSING SUPPORT SYSTEM
BENEFIT-COST ANALYSIS

TABLE OF CONTENTS

Page

TABLE OF CONTENTS
(continued)

Preface

This is the fourth and last in a series of four reports on the Licensing Support System (LSS) prepared by the DOE Office of Civilian Radioactive Waste Management (OCRWM) for the Office of Management and Budget (OMB). The LSS is an information management system intended to support the needs of all the parties involved in repository licensing, including the Department of Energy (DOE) and the Nuclear Regulatory Commission (NRC). These reports are:

Preliminary Needs Analysis

Preliminary Data Scope Analysis

Conceptual Design Analysis

Benefit-Cost Analysis

The Preliminary Needs Analysis, issued in February 1988, and the Preliminary Data Scope Analysis, issued in March 1988, constitute the system requirements basis for developing a Conceptual Design, which was issued in May 1988. The Benefit-Cost Analysis presented in this report evaluates alternatives within this conceptual design. These four reports, and subsequent refinements, are intended to provide the basis for determining the LSS design specifications.

EXECUTIVE SUMMARY

A Benefit-Cost Analysis has been performed on the LSS design alternatives which were identified in the Conceptual Design Analysis (DOE, 1988c). The purpose of the analysis is to compare these alternatives and derive conclusions leading toward a preferred design. The analysis does not extend to estimating the costs or benefits of attempting to achieve the licensing decision without an LSS, since such a system is expected to be required by 10CFR2, it does include an analysis which estimates the costs associated with a delay in the operation of the repository at $195 million per year.

The alternatives examined include a Base Conceptual Design and seven variants. The Base Conceptual Design has the following major features:

1) Headers and searchable full text of all documents suitable for inclusion in LSS

2) Bit-map images of all documents in LSS
   - reproduction of documents for quick distribution from central location
   - on-line display and local printing at special workstations

3) Centralized search system and on-line optical disk image system in Washington, DC or Las Vegas, NV

4) Multiple capture systems for:
   - scanning
   - text conversion
   - correction
   - cataloging

5) Workstations capable of displaying readers, ASCII text and images

6) Support for workstations displaying headers and ASCII text only

7) Retrieval through structured index searching of cataloging information and software full-text searching of documents

8) Electronic mail

The Base Conceptual Design hardware architecture is shown in Figure 1 and the corresponding communications architecture is given in Figure 2.

The variants examined differed from the Base Conceptual Design in the following ways:

I. Two search and image systems replicating the data base, rather than one, located in Washington, DC and Las Vegas

II. Hardware full-text search, rather than software

III. No workstations capable of displaying images

IV. Microform digitization rather than optical disk storage of images

V. Off-line microform printing rather than on-line bit-mapped image system

VI. Re-keying text rather than text conversion from scanned bit-map

VII. Combination of III, V and VI above.

Table 1 summarizes the features of each of these alternative systems, in terms of full-text search method, local image availability, image storage method, hardcopy distribution method and text capture process.

The cost analysis was performed over a 10 year life cycle of the LSS beginning with system design in FY89 and extending through 8 years of operation, since the initial hardware is expected to be suitable for replacement in 10 years. Life cycle costs include data capture, system design, system procurement, and system operation. The Base Conceptual Design and all seven variants resulted in similar life cycle costs varying from $192 million to $236 million in 1988 dollars. The following conclusions can be drawn:

1) The predominant factor in the total cost was associated with data capture (the process of collecting and preparing the information for loading into the system), accounting for approximately 62 percent of the total costs.

2) With the exception of Variant I, there is only a small (less than 10%) difference in life cycle costs in the variants compared with the base. Four variants differ only 2% or less from the base. This is due primarily to the fact that much of the total cost is associated with the capture process, and the variants do not impact this process significantly.

3) Only one alternative (Variant V) is lower in cost than the base.

4) The data capture costs (and therefore total life cycle costs) are not very sensitive to the rate at which pages are processed but are primarily dependent upon the total number of pages. Increasing the rate at which the backlog is processed from 18,000 pages/day to 20,000 pages/day in the period FY91 to FY94 would result in the backlog being loaded one year earlier at an increase in total cost of only about $1 million.

5) The LSS costs are primarily labor intensive. Using the figures for the Base Conceptual Design, the major costs contributors are:

Labor	70%
Hardware	16%
Facility	7%
Telecommunications ...	4%
Hardcopy production ..	3%
Software	2%

A benefit comparison of the alternative designs relative to the base was performed to provide some measure of effectiveness of certain non-quantifiable attributes (at least in dollar terms) of the systems. (It should be noted that quantifiable "benefits" such as lower operating costs were included in the cost calculations rather than as benefits.)

The numerical values obtained to quantify benefits must be considered in light of the parameters used in the benefits analysis. The parameters were selected with the intent of highlighting the benefits associated with distinctions between the Base Conceptual Design and variants. Overall system performance, as measured by the search criteria, is constant across most of the systems, and serves to keep the distinctions in perspective. Some general conclusions from the analysis are:

1) The single biggest factor in increasing user benefits is the lowered load factor in Variant I: the more computer power available to each user, the better. Note that if the total computing capacity is kept equal to the Base Conceptual Design (i.e., the low benefit load factor sensitivity for Variant I), then benefit decreases to somewhat less than base.

2) The use of off-line image storage and retrieval significantly reduces relative benefit unless the average time to receive a print request is reduced to one day; in that case, the impact is small.

3) Over the ranges examined, Capture Delay and Image Quality have only minimal impact on relative benefit in this analysis. Re-keying text (Variants VI and VII) is comparably useful to the Base Conceptual Design, in spite of the longer Capture Delay. Off-line microform storage of images (Variants V and VII) is appreciably less useful because of Image Return Time, not Image Quality.

As a result of both the cost and benefit analyses, the Base Conceptual Design appeared to be the overall preferred design; however, the results for some other variants were sufficiently close that some design alternatives must continue to be considered.

Table 18 below summarizes, in relative terms, the results of the cost and benefit analyses among the alternative configurations:

TABLE 18. COST AND BENEFIT STUDY

Alternative	Relative Cost	Relative Benefit
Base	1.00	1.00
Variant I	1.21	1.11
Variant II	1.01	0.90
Variant III	1.01	0.85
Variant IV	1.02	0.95
Variant V	0.98	0.76
Variant VI	1.08	0.96
Variant VII	1.06	0.75

A sensitivity analysis on data volume (number of pages), percent of data in searchable full text, and number of simultaneous users did not indicate any anomalies in the conclusions reached from the analyses of costs and benefits. It did point out, however, the sensitivity of the total life cycle costs to data volume and to a lesser extent to the percent of volume placed in searchable full text. To force a cost reduction in the LSS through a reduction in either data volume or percent full text would risk a loss of system usefulness and user confidence, which could jeopardize the basic goal of shortening the repository licensing process.

1.0 INTRODUCTION

This report provides an analysis of the benefits and costs associated with design concepts of the Licensing Support System (LSS) developed from the results of a systematic analysis of needs and requirements as documented in the three previous reports in this series.

1.1 Purpose and Scope

A Benefit-Cost Analysis is a systematic approach for comparing alternative methods to satisfy an objective. In this case the objective is to provide an electronic information management system which contains the information relating to the licensing of a geologic repository for the disposal of high-level radioactive waste. The major steps necessary to performing this analysis are:

1) Identifying the system requirements
2) Identifying alternative conceptual designs
3) Analyzing the costs and benefits of the alternatives.

The Preliminary Needs Analysis (DOE, 1988a) and the Preliminary Data Scope Analysis (DOE, 1988b) document the results of the first step, and the Conceptual Design Analysis (DOE, 1988c) documents the result of the second. This report compares the costs and benefits of the alternatives.

Since the LSS is expected to be required by regulation (NRC, 1987) and since in any event it is considered necessary to comply with the statutory requirement that a decision is reached on issuance of a repository construction authorization within three years, the scope of this analysis is limited to various alternative methods of satisfying the needs for the system. It does not extend to estimating the costs (or benefits) of attempting to achieve the licensing decision without such a system.

Nevertheless, it is probable that without an LSS, the licensing process would be extended, both in the review of the application for construction as well as for the application for operation. These extensions would linearly extend the date of operation for the repository. The costs associated with extending the program and with continued (and increased) storage of fuel at the reactor site have been estimated and included for information.

1.2 Background

As noted in the previous reports in this series, several studies were performed to identify the needs and requirements of the LSS and a few previously documented studies exist on estimating the size of the data base to be incorporated. The study produced by Arthur Young for DOE (DOE,1987) carried the process to the extent of examining various alternative conceptual architectures and even included some cost estimates. Since that time, however, the LSS requirements (especially those derived from the Negotiated Rulemaking process) have become much better defined, permitting a detailed analysis of system design over a narrower range of functionality.

2.0 APPROACH

Most benefit-cost analyses of computer-based information management systems are a comparison of various alternative changes or improvements with the present method of achieving objectives. In this case, benefits and costs are more readily definable and even the decision on assigning various quantifiable aspects to either costs or benefits is more apparent. Typically one allocates the cost of replacing the present system with alternative "A" as a cost, and if that alternative were to exhibit lower maintenance costs, for example, the operational savings would be quantified as a benefit.

In the case of the LSS, the picture is not so simple. There is no manual or automated precedent for the operation of a central management system for the information relating to licensing of the repository. Indeed there is really no precedent for this particular licensing process itself. Therefore the approach to be followed in this analysis is probably as unique as the situation, but it is nevertheless both sufficiently rigorous and comprehensive to provide a rational basis for choosing among alternative concepts.

The first step was an analysis of costs for each alternative considered. The alternatives were the Base Conceptual Design and the seven variants of that design as described in the Conceptual Design Analysis and summarized in the next section. The base design is neither a present case or a preferred design. It is simply the alternative which appeared to best meet the defined needs and at the same time represented a low risk technical solution. Since this was clearly not a point of reference for the calculation of benefits associated with the variants, it was decided to maintain all quantifiable benefits as variations in the costs. If one alternative results in lower maintenance costs, this is reflected in a lower life-cycle cost for that system rather than a benefit over the Base Conceptual Design. The cost calculations are therefore performed for each case (base and seven variants) as total life-cycle costs, including operation, maintenance, data capture, etc.

The cost analysis has been performed over a 10 year life-cycle period beginning with fiscal year 1989 and continuing through fiscal year 1998. The starting period was chosen to cover the period of major design and procurement. The ending period was chosen to extend over the first major milestones in the licensing process, i.e. submission of the application and issuance of the construction authorization. As noted in the high-level waste (HLW) repository program time-line (Figure 1 of the Preliminary Data Scope Analysis) the LSS usage is expected to significantly decrease in 1998 and would therefore represent a good opportunity for system replacement. The cost analysis therefore assumes the full life-cycle cost of a single system, having no residual value at one end of its life and incorporating no costs for follow-on design or replacement.

All of the alternative designs meet all the basic requirements which have been identified in the previous analyses. An analysis of life-cycle costs, which implicitly includes benefits associated with lower costs, would therefore lead to the conclusion that the lowest cost alternative would be the preferred design. However, the alternatives do not meet all of the requirements in the same way or to the same degree. For this reason there must be some methodology to compare these alternatives in the manner in which they perform relative to the requirements. This analysis is performed as the benefits section of the report (Section 5). Although these aspects are not quantified in absolute terms, they are rigorously defined in relative quantities (weights) and combined to provide a relative ranking of alternatives.

The accuracy of the analysis is subject to the accuracy of the many assumptions which must be made in order to quantify the costs. These assumptions are listed in Section 4 and the various tables of Appendix B as they are incorporated. In almost all cases the assumptions are either reasonably well known quantities, have a small effect on the total cost, or equally affect all of the alternative costs such that variations in these assumed quantities would not affect the conclusions of the study. In some cases, however, the assumptions could have significant effects on the calculations, and in those cases, variations in those figures were made as a sensitivity analysis. The three variables modified in the sensitivity study included:

1) Data volume (number of pages included in the system)
2) Percent of the data in "full-text"
3) Number of simultaneous users.

The results of this study are presented in Section 7.

This process results in a sufficiently rigorous analysis to provide a basis for a reasonable choice among the alternative systems. It does not in itself provide any basis for determining the viability of the LSS itself, since, as noted in Section 1, the system is considered to be a vital component in attempting to meet the Congressionally mandated 3-year period for a decision on the construction authorization.

As a mechanism to compare the magnitude of costs associated with construction and operation of the LSS, however, one can contemplate that without the LSS, the repository operation would be delayed for some time due to the lengthened licensing process. While the extent of the delay could only be conjectured, it is generally agreed that it would be a period of years as opposed to days or months. Costs associated with a repository operation delay can be primarily allocated to two categories: development and evaluation costs (D&E) of the program and at-reactor fuel storage costs. The total of these are estimated to be approximately $195 million per year in 1988 dollars. Details of this calculation are found in Appendix A.

3.0 DESCRIPTION OF ALTERNATIVES

The Conceptual Design Analysis provides the detailed description of the design concepts for the Base Conceptual Design and seven design variants. This section summarizes these descriptions of those alternatives in order to provide a frame of reference for the cost and benefit analyses presented in Sections 4 and 5.

3.1 Base Conceptual Design

The LSS is to provide on-line access to the information required to support licensing the construction and operation of a nuclear waste geologic repository, as specified by the negotiated rulemaking process. The access will consist of on-line search capability of fielded data (referred to as headers) and full text of each document in the system. The headers, text and its images can be viewed on-line *via* special workstations. The Base Conceptual Design, illustrated in Figure 1, includes the hardware, software and operations necessary to capture the information, load it into the system, store it, provide on-line query and display, and to distribute hardcopy upon request. The LSS concept presented in the Conceptual Design Analysis report is comprised of the capture system, search system, image system, communications and workstations, each of which is summarized in the following sections.

3.1.1 Capture System

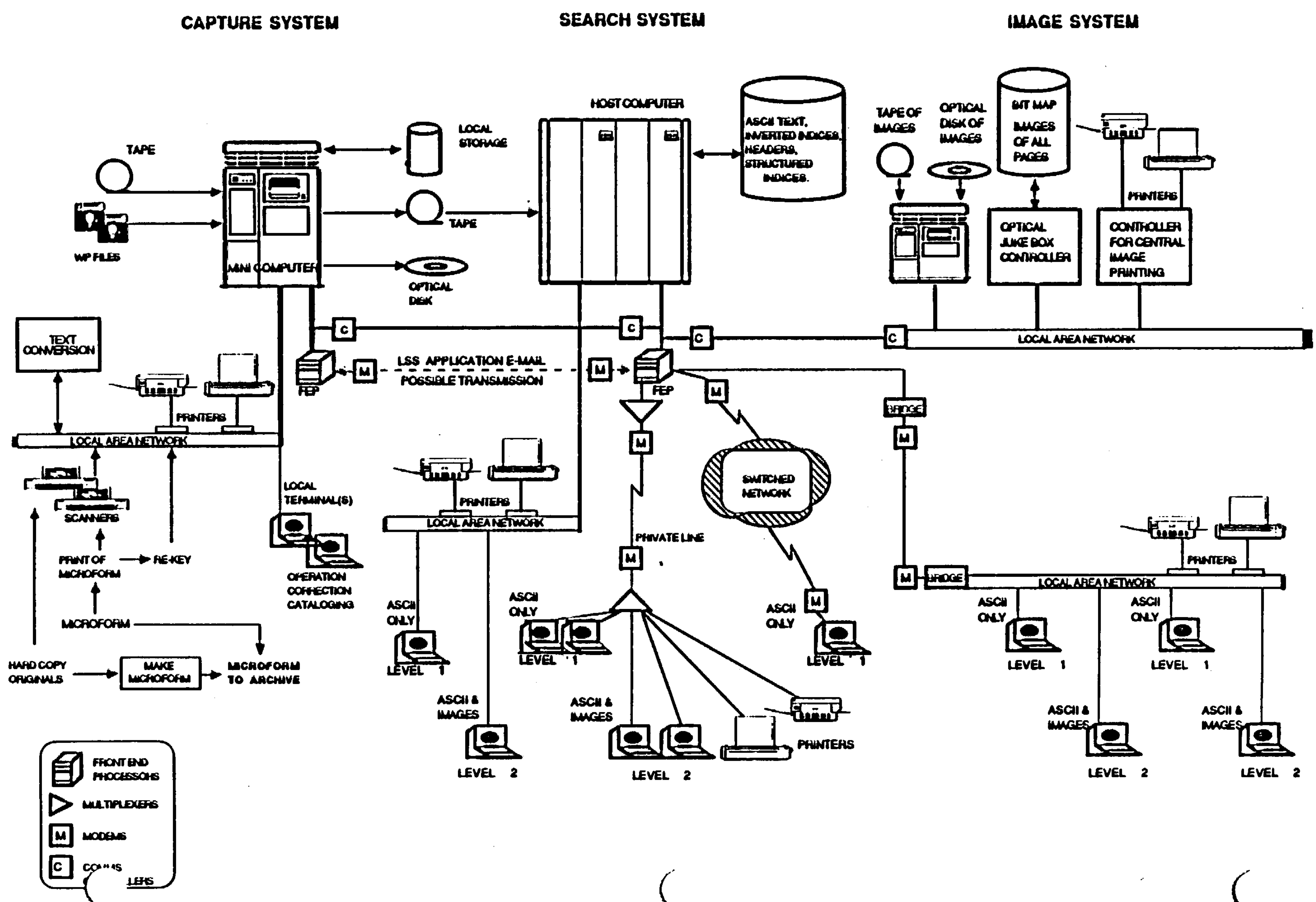
The capture system consists of the operating procedures, computer, image and character recognition equipment, software and facility to process documents submitted for entry into the LSS. The capture station processes are:

Document receipt, duplication check and accession number assignment
Scan to create bit-mapped image
Microfilm for archive
Obtain ASCII text from the bit-mapped image
First quality check
Creation of the header
Final quality check

In order to handle the expected load, the capture system must consist of six stations, each with the capacity to process 3,000 pages per day, operating two shifts for a total of 15 hours per day. The six stations will be at three locations in Las Vegas and Washington, DC. The final distribution depends on the distribution of new and backlog material to be entered.

FIGURE I

BASE CONCEPTUAL DESIGN HARDWARE ARCHITECTURE



3.1.2 Search System

The search system is the operating procedures, computer equipment and magnetic storage devices, software and facilities which load the information prepared by the capture process and make it available for on-line query and retrieval. The search system also supports electronic mail and access to the LSS Regulations Access Subsystem.

The Base Conceptual Design architecture consists of a single search system, co-located with the image system described below. It utilizes a full-text data base management system (DBMS) to prepare structured and inverted indices and store the indices, headers and full ASCII text. LSS software in conjunction with the DBMS provides the on-line access and interfaces with the image system. The search system computer can be either a large mainframe or a tightly coupled cluster of super-minicomputers.

3.1.3 Image System

In the Base Conceptual Design the image system stores images of all documents in the LSS on optical disks for on-line retrieval and display on Level 2 workstations or for off-line volume printing of documents via high speed laser printers. It is connected directly to the search system from which it receives commands. Output is routed directly to the workstations via the communications system.

The image system consists of three components which are interconnected by a local area network. The components are: (1) image preprocessor; (2) optical disk jukebox controller and jukebox storage unit(s); and (3) printer controller and high speed laser printer(s).

3.1.4 Communications System

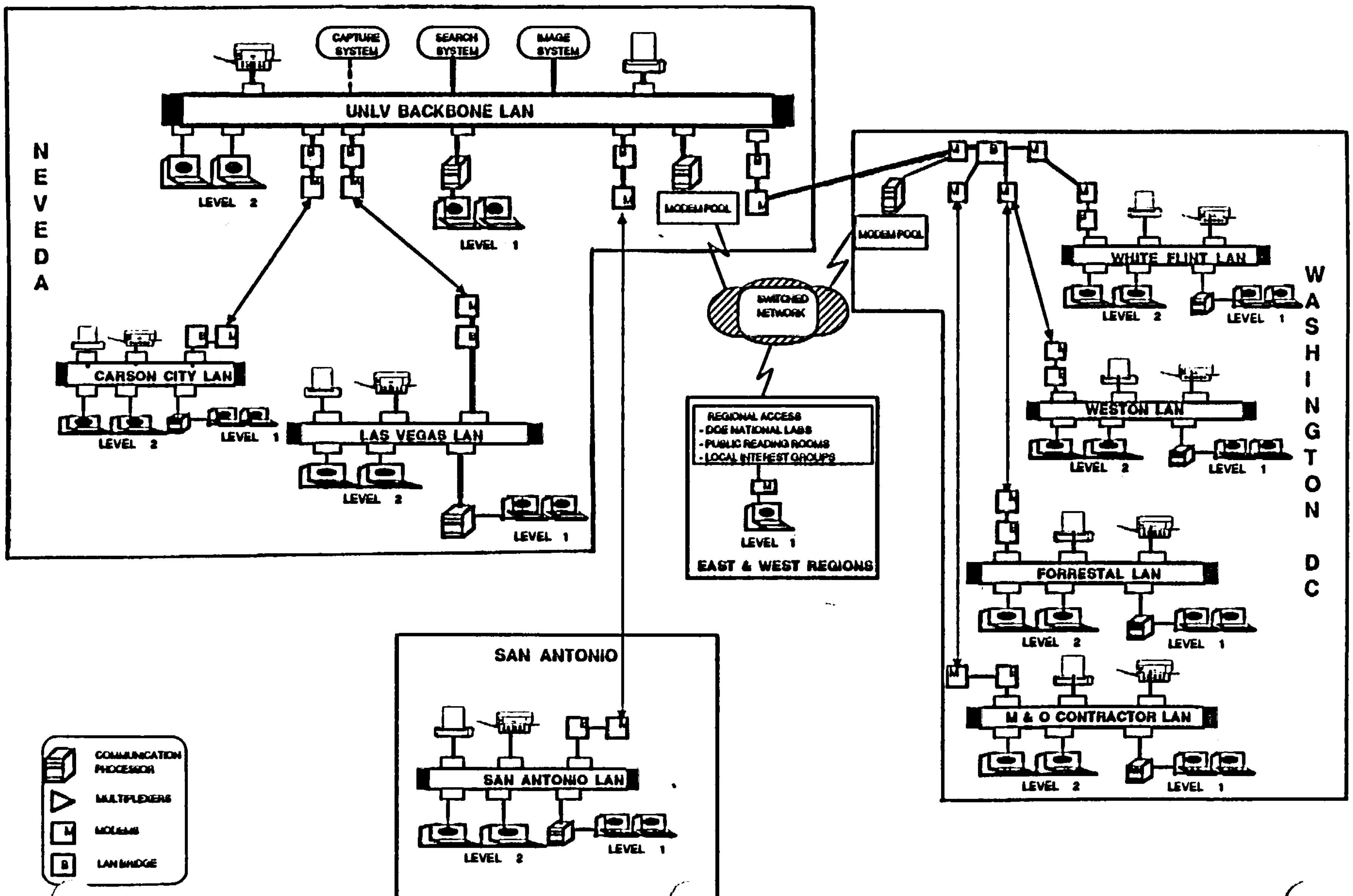
User queries and display requests, electronic mail messages, and ASCII text and image data to be displayed or printed at the workstations are transmitted over the LSS communications system.

Every LSS user, regardless of location, will view the LSS as though directly connected to the search system. The communication system consists of five technologies: local area networks (LANs), intelligent bridges, high speed multiplexers, high speed modems for voice grade switched circuits, and intelligent communications processors. Figure 2 shows the network topology with Las Vegas as the location of the search and image systems. Users in the Washington, DC area, Nevada and Texas are supported by LANs while other users are supported through dial-up services.

3.1.5 Workstations

Users access the LSS through workstations connected to the LSS communication system. There are two types of workstations supported, Level 1 and Level 2. Level 2 workstations can display full page images and full page ASCII text. They also have local image print capability. The Level 1

FIGURE 2. **LSS TOPOLOGICAL COMMUNICATION NETWORK ARCHITECTURE**



workstations have no capability for on-line display of images and can display only pages of ASCII text, part of a page at a time.

3.2 Variants to Base Conceptual Design

Several major features of the LSS may be implemented in more than one viable way. This section summarizes seven variants on the Base Conceptual Design which present such alternatives. Although all variants satisfy the functional requirements to some degree, the efficiency and manner in which they are satisfied vary. The degree of satisfaction is measured by the benefit analysis discussed in Section 5. Table 1 provides a summary of the principal features of the variants.

3.2.1 Variant I - Full Replicated Nodes

Variant I differs from the Base Conceptual Design in that it has two fully replicated search and image system nodes, one located in Washington, DC and one in Las Vegas, NV. The system architecture consists of the identical hardware and software configuration as the Base Conceptual Design. This variant was chosen to exploit the data security that full redundancy offers for data base backup and recovery. Additional advantages include greater computing capacity available to each user since each node has about half of the total users, higher overall availability if the centers are linked plus potentially lower recurring communication charges, since the total leased lines capacity required would be less.

3.2.2 Variant II - Hardware Full-Text Search

In the Base Conceptual Design full-text search is implemented via storage of the full text plus creation of an inverted index. The full-text data base software uses both to respond to user queries. Variant II replaces them and the text data base manager with specialized hardware processors. The hardware compresses the full text and stores it on very high transfer rate disk drives for subsequent searching. The search is performed serially through all of the compressed full text. The search speed is a function of data base size, disk transfer speed, and the relative size of the query relative to the width of a hardware comparater. The hardware full-text processor is directly attached to the search host computer from which it receives search requests and returns full text. Multiple hardware processors can be connected to the host computer. All other functions (E-mail, header searches, etc.) performed by the search system host computer remain the same as the Base Conceptual Design.

3.2.3 Variant III - Images Are Not Supported At Workstations

In this variant from the Base Conceptual Design the capability to view electronic (bit-mapped) images on the screen at the Level 2 workstations is excluded. This variant was selected since this capability was not identified as necessary by all potential LSS users in the Preliminary Needs

TABLE 1. SUMMARY OF LSS DESIGN VARIANTS

SYSTEM VARIABLES

Variant	Number of Search/Image Nodes	Full-Text Search System	Local Images	Image Storage Media	Hardcopy Distribution System	Full-Text Capture Process
Base	1	Software/DBMS	at Level 2 Workstations	Optical Disk	Hardcopy from Optical Disk	OCR scanned bit-mapped image
I	2	Software/DBMS	at Level 2 Workstations	Optical Disk	Hardcopy from Optical Disk	OCR scanned bit-mapped image
II	1	Hardware	at Level 2 Workstations	Optical Disk	Hardcopy from Optical Disk	OCR scanned bit-mapped image
III	1	Software/DBMS	None at Workstations	Optical Disk	Hardcopy from Optical Disk	OCR scanned bit-mapped image
IV	1	Software/DBMS	at Level 2 Workstations digitized from microfilm	Microfilm	Hardcopy from digitized microfilm	Bit-mapped image captured from microfilm
V	1	Software/DBMS	None at Workstations	Microfilm	Hardcopy from microfilm (like NTIS service)	OCR scanned bit-mapped images
VI	1	Software/DBMS	at Level 2 Workstations	Optical Disk	Hardcopy from Optical Disk	Re-keying of all text submitted as hardcopy
VII	1	Software/DBMS	None at Workstations	Microfilm	Hardcopy from microfilm (like NTIS service)	Re-keying of all text submitted as hardcopy

Analysis. To compensate for not being able to display images at a workstation, the requests for printing images is expected to be appreciably greater.

3.2.4 Variant IV - Microform Digitizers in Capture and Image Systems

In this variant on the Base Conceptual Design the changes occur in the capture and image systems. Equipment capable of creating digitized (bit-map) images from microform is used to create the OCR input for documents available only on microform. Microform replaces optical disk as the medium for the storage of images. This variant was chosen to reflect the availability of automated microform systems and the possibility that some portion of backlog documents are available only on microfilm.

3.2.5 Variant V - Microform Off-Line Image Storage and Retrieval

Variant V replaces the on-line image system with a off-line service for obtaining hardcopy or microform copies of LSS documents. This is similar to the way commercial and existing DOE bibliographic data base services provide document copies to their users. For example, DIALOG allows users to order documents from NTIS as a command after locating the document.

This variant was developed to present a low-tech solution to meeting the hardcopy receipt time requirements (of 2 to 3 days) identified in the Preliminary Needs Analysis. The capability to view electronic (bit-mapped) images on the screen at the Level 2 workstations is not supported, as in Variant III.

3.2.6 Variant VI - Full Text *via* Re-keying

In this variant, there is no automated process to create searchable text (OCR). The conversion of hardcopy text to ASCII is accomplished by re-keying the document. An expected 99.8% accuracy of data by re-keying would be achieved by double keying the original source document. Since the text conversion will be accomplished by re-keying there will be no requirement for optical character recognition equipment, and associated software. However, the re-keyed documents would require processing through a digital scanning device since bit-mapped image capture and storage is required.

3.2.7 Variant VII - Combined Variants III, V and VI

Variant VII combines hardware and software changes for the Level 2 workstations (Variant III), the removal of the on-line image system (Variant V) and the re-keying of all documents instead of OCR (Variant VI). This variant was created to present a conceptual design with the lowest schedule risk which minimally meets the requirements presented in the Preliminary Needs Analysis.

4.0 COST ANALYSIS

4.1 Cost Calculations

This section describes the method used to develop the estimates for the parameters required by the cost model to compute the total life-cycle cost of the development and operation of the LSS over 10 years. (It is expected that the initial LSS hardware/software environment will be obsolete and require replacement at that time.) These parameters include the year-by-year input receipt rates; number of personnel needed by skill group; labor rates for each skill group; capture system, search system, image system and communications hardware purchase costs; software purchase and development costs; facility preparation and operation costs; and hardcopy distribution costs. Labor and operation costs are based upon industry standards, and hardware costs are not vendor or configuration specific.

The full cost model for the Base Conceptual Design is included in Appendix B for reference.

4.1.1 Workload Processing

Recognizing the dependency of the cost calculations on the type and number of pages to be processed, the initial task in the cost analysis was to determine the number of pages and documents which would be entered into the LSS each year, and the processing required for ASCII conversion. The basis for the calculations was the estimates of pages identified in Table 8 of the Preliminary Data Scope Analysis. This table (as corrected in the Conceptual Design Analysis; DOE, 1988c) provided both a high and low estimate of the number of pages which are candidate for inclusion in the LSS during the years 1990 through 2009, and of the backlog which would be accumulated by 1990.

As the first step in the process, the values were converted from calendar to fiscal year basis, since all costs are calculated for fiscal years. For the high estimate of data volume, this resulted in a cumulative figure of 27.6 million pages appropriate to be in the system by the end of the analysis period (September 1998). The second step was to determine the schedule in which the backlog would be processed and loaded into the system. The schedule constraints used are:

1) The capture process will begin in January 1990
2) The backlog must be loaded by October 1994, 6 months prior to the estimated date of license application.

During the period FY90 through FY94, it is assumed that all the new material appropriate for LSS generated during each year will be processed as well as some percentage of the backlog. (This backlog at the beginning of FY90 is estimated to exceed 10 million pages.) The percentage of backlog processed in a given year was adjusted to provide a relatively even workload over the years FY91 to FY94 with a lower processing rate for FY90 (to allow for a "ramping-up" period). The resultant processing rates were slightly over 10,000 pages per day for FY90, and close to 18,000 pages per day for FY91 to FY94. Following this peak effort, the loading rate drops to the rate required to process the new material produced during each year, approximately 10,000 pages per day in FY95 decreasing to 6,000 pages per day in FY98 (cf. Table B.4 of Appendix B).

The final step in the process was to provide an estimate for the distribution of the workload for the number of pages of ASCII text to be converted, indexed, and stored. Consistent with the estimates in the Data Scope Analysis, it was assumed that 95% of the pages are textual; the remaining 5% being figures, graphs, or other non-textual material. Among the pages to be converted to ASCII, certain documents will be available in word processing form (i.e. the text was generated by electronic means and an ASCII form of the text was submitted). The amount of input in this form will increase progressively with time. It was assumed that 5% of the pages received in FY89 would be in word processing form, increasing to a maximum of 75% in FY93 and later. All other pages require conversion by optical character recognition (OCR) devices for the Base Conceptual Design, with the exception of 5% of the pages generated prior to FY89 which were assumed to require re-keying (due to their lack of acceptability for OCR processing) (cf. Table B.5 of Appendix B). These assumptions, when applied to the daily processing rate, resulted in an estimate of pages to be processed by each method of ASCII conversion (Table B.6 of Appendix B).

4.1.2 Capture Process

The cost for the capture process was derived from a process flow model developed for the definition of the Base Conceptual Design and the variants presented in the Conceptual Design Analysis. The refinement of this model involved a detailed analysis of the data conversion effort required to support the estimated workload. This analysis resulted in the definition of the required capture station processes as listed in Section 3.1.1 of the Conceptual Design Report. These processes were used as the basis for estimating the labor requirements (numbers of people and skill levels) and the hardware requirements (generic technical specifications and numbers of units and/or systems) for the capture operation.

The process model for the Base Conceptual Design and the corresponding labor and hardware requirements was analyzed and appropriately modified for each design variant.

Labor costs for the data capture operation were estimated based on industry salaries for the required skills, and a salary burden factor of 2.0 was applied.

Hardware cost were estimated based on published price lists of vendor products which met the generic technical specifications.

Facility costs are representative of industry standards and include both non-recurring and recurring costs. Non-recurring (start-up) costs include such preparation costs as raised flooring, special power and air conditioning, and fire and electrical systems. Recurring costs include lease, utilities, communications, freight, maintenance, operational supplies, and furniture and equipment rental.

The software development costs to support the LSS capture process were estimated based on the process flow model described above. This model served as the basis for further definitization of the software requirements from the base design presented in the Conceptual Design Analysis report. The major modules to be developed were identified as follows:

- Cataloging header data base
- Duplicate check data base
- Image process control
- Text process control
- E-Mail interface
- Quality control
- Production control
- Systems Administration

The characteristics of each major module were specified by providing an estimate of the number of the following items each will have:

- Data elements
- Data Base records
- Screens - both input and query
- Reports - both on-screen and printed
- Processes

A model was used to compute the total design and development hours required for each major module based on the characteristics of each. The model extended the total hours to get labor dollars and computer support dollars. The software development required for the Base Conceptual Design was computed using the model and estimates of the number of each of the module items. The software development costs for each of the variants was computed by estimating how much of each major module was required for each variant. That derived percentage was then applied to the number of hours needed for the base design for that major module. The total software development cost for each variant was extended using the total of the major module hours. Labor costs for data capture software development were estimated based on industry salaries and burden rates for the required skills.

4.1.3 Search/Image System

The search system cost estimate was based on a computer model developed to estimate the CPU requirements in Millions of Instructions per Second (MIPS). This model was based on query complexity, data base size and the number of simultaneous users. Using the results of this model, the estimated size of the search system processor for the Base Conceptual Design and for each variant were made. Cost estimates were based on published vendor price lists for state-of-the-art products which met the processing power and input/output (I/O) rates estimated by the computer model.

The estimated cost for magnetic disk storage was based upon a disk system with an individual disk capacity of 7.5 Gbytes and a disk controller with a capability for four independent paths and 128 Mbytes of cache memory. The number of disk controllers and the number of drives per controller were configured to provide minimum device contentions and maximum utilization of I/O channels and transfer rates. Cost estimates were based on published price lists for vendor products which met the disk capacity and transfer rates estimated by the computer model.

The image system cost estimate was based on a market/technical analysis of available image systems, including both electronic systems based on optical disks and microfilm systems as required by some of the variants. Representative, non-vendor-specific systems which met the requirements for the base design and variants were chosen as a basis for forming the hardware cost estimates.

The search/image system software development costs were determined using the methodology described in the capture system software development (Section 4.1.2). The major modules for the search/image system are:

- Header data base
- Full text data base
- Query menus
- Query screens
- Prompt dialog user interface
- On-line help
- Image display
- Query and results save and reuse
- Multiple partition searching
- Sample Inventory Management System interface
- Data base load manager
- Performance monitor
- Systems administration

Labor costs for operation of the search/image system and facility were based on an analysis of staffing for similar operations industry wide. Cost estimates include the premise of continuous four shift operation. Labor costs were estimated based on industry salary rates, burdened at a factor of 2.0. Facility costs include the same factors as the non-recurring and recurring costs for the capture process.

Estimated costs for the user workstations, both hardware and software were included only for the Level 2 workstations (capable of on-line electronic image display). The Level 1 workstation is to be a personal computer-based system which will be supplied by the user.

Communication costs include initial equipment investment, initial telephone circuit installations, and recurring telephone circuit costs (monthly). To estimate the initial equipment investment, all LSS locations were divided into categories based on the number and type of users that were estimated for each site. The equipment requirements for each category of location were specified and cost estimates were given for the number of terminal connections necessary for each site. The number of sites per category and the total initial equipment investment were calculated using a customized spreadsheet for all LSS locations for the base and each variant of the conceptual design.

The initial telephone circuit installation and recurring circuit costs were estimated using a telecommunication network modeling tool. The tool offers the tariffed rates for communication services from the various communication Common Carriers (such as AT&T). The tariffs vary based on the rate of speed (for communication), the distance between the communication end points and the types of facilities and services desired (for example, conditioned lines for improved quality). The most sensitive cost variable is the rate of speed for the communication. To estimate this value, the number of users at a specific site or in a region and the amount of information the users are expected to send and receive are considered along with the cost to connect and maintain a connection between the LSS and a Public Data Network Service (such as Telenet) to facilitate users on terminals that were not dedicated to LSS usage (i.e., dial-up users). These factors were analyzed for the base and each variant of the conceptual design.

All design alternatives include a requirement to provide hardcopy of LSS records upon receiving a request from a user. The system to provide the copies is included in the image system and incorporates printing from electronic or microform images as appropriate for the alternative design. A detailed print load analysis was performed which estimated the number of copies which would be required based on estimates of number of users, number of sessions per user, number of queries, and the number of records which meet a query. Additional calculations based on National Technical Information Service (NTIS) experience and expected information requirements of the users were made to verify the results. Appendix C provides background on these estimates. The resulting figures were weighted to the number of pages to be generated each year in order to reflect both a dependency on the size of the data base and program activity. For the base design, hardcopy production varies between a low of 5 million pages per year in FY91 to a high of 10 million pages per year in FY94 and FY95. A cost of 10 cents a page is included in the cost calculations to account for supplies and shipping costs, assuming that expedited shipping will be required in most cases.

4.2 Results

The total 10-year life-cycle costs for the Base Conceptual Design and the seven variants are presented in Tables 2 through 9, and summarized as follows:

Alternative	Life-Cycle Costs (1988 dollars)
Base	$ 195 million
Variant I (2 sites)	$ 236 million
Variant II (Hardware full text)	$ 197 million
Variant III (No on-line images)	$ 196 million
Variant IV (Microform)	$ 198 million
Variant V (No on-line images, Microform)	$ 192 million
Variant VI (Re-key)	$ 210 million
Variant VII (Combination of III, V and VI)	$ 207 million

Costs in Tables 2 through 9 are detailed by fiscal year and grouped as non-recurring or recurring, as well as according to their association with the capture process or the storage and retrieval process (search and image systems). The definition of the cost elements are:

	Capture	Search/Image
Non-recurring	Hardware procurement Software procurement Software development Facility preparation	Hardware procurement Software procurement Software development Facility preparation
Recurring	Data prep labor System operating labor Facility operation Maintenance Subcontract services	System operating labor Facility operation Maintenance Telecommunications Hardcopy production

In reviewing the figures presented in these Tables, several points are apparent:

1) The major cost of the LSS is associated with data capture.

2) With the exception of Variant I, there is only a small (less than 10%) difference in life-cycle costs in the variants compared with the base. Four variants differ only 2% or less from the base. This is due primarily to the fact that much of the total cost is associated with the capture process, and the variants do not impact this process significantly.

3) Only one alternative (Variant V) is lower in cost than the base.

Some additional observations come to light with a more detailed review of the results.

1) The data capture costs (and therefore total life-cycle costs) are not very sensitive to the rate at which pages are processed but are primarily dependent upon the total number of pages.

Table 2. BASE CONCEPTUAL DESIGN LIFE CYCLE COST BASIS
(IN THOUSANDS OF DOLLARS)

	FY89	FY90	FY91	FY92	FY93	FY94	FY95	FY96	FY97	FY98	TOTAL
NON-RECURRING COSTS											
Capture	5,556	1,614	0	0	0	0	0	0	0	0	7,170
Search/Image	1,346	9,718	3,519	2,917	2,072	1,839	0	0	0	0	21,412
SUBTOTAL	6,902	11,332	3,519	2,917	2,072	1,839	0	0	0	0	28,581
RECURRING COSTS											
Capture	394	11,641	19,438	18,883	17,629	15,629	8,878	8,094	7,161	6,165	113,911
Search/Image	0	0	5,555	6,076	6,506	6,882	7,053	6,944	6,815	6,643	52,474
SUBTOTAL	394	11,641	24,993	24,959	24,135	22,511	15,930	15,037	13,976	12,808	166,385
TOTAL COSTS	7,296	22,973	28,512	27,877	26,207	24,350	15,930	15,037	13,976	12,808	194,966
								Present value @		3.0%	167,662
										10.0%	122,153

Table 3. VARIANT I - LIFE CYCLE COST BASIS
(IN THOUSANDS OF DOLLARS)

	FY89	FY90	FY91	FY92	FY93	FY94	FY95	FY96	FY97	FY98	TOTAL
NON-RECURRING COSTS											
Capture	5,556	1,614	0	0	0	0	0	0	0	0	7,170
Search/Image	1,346	14,820	5,019	5,602	3,911	3,678	0	0	0	0	34,375
SUBTOTAL	6,902	16,434	5,019	5,602	3,911	3,678	0	0	0	0	41,544
RECURRING COSTS											
Capture	394	11,641	19,438	18,883	17,629	15,629	8,878	8,094	7,161	6,165	113,911
Search/Image	0	0	8,485	9,157	9,854	10,415	10,769	10,660	10,532	10,360	80,231
SUBTOTAL	394	11,641	27,923	28,039	27,484	26,043	19,646	18,754	17,692	16,524	194,142
TOTAL COSTS	7,296	28,075	32,942	33,641	31,394	29,721	19,646	18,754	17,692	16,524	235,686
								Present value @		3.0%	202,189
										10.0%	146,537

Table 4. VARIANT II - LIFE CYCLE COST BASIS
(IN THOUSANDS OF DOLLARS)

	FY89	FY90	FY91	FY92	FY93	FY94	FY95	FY96	FY97	FY98	TOTAL
NON-RECURRING COSTS											
Capture	5,622	1,614	0	0	0	0	0	0	0	0	7,236
Search/Image	1,793	11,750	6,319	1,568	723	490	0	0	0	0	22,643
SUBTOTAL	7,415	13,364	6,319	1,568	723	490	0	0	0	0	29,878
RECURRING COSTS											
Capture	394	11,641	19,438	18,883	17,629	15,629	8,878	8,094	7,161	6,165	113,911
Search/Image	0	0	5,691	6,492	6,787	7,028	7,064	6,955	6,827	6,655	53,499
SUBTOTAL	394	11,641	25,129	25,375	24,416	22,657	15,941	15,049	13,987	12,819	167,410
TOTAL COSTS	7,809	25,005	31,448	26,943	25,139	23,147	15,941	15,049	13,987	12,819	197,288

Present value @	3.0%	170,040
	10.0%	124,545

Table 5. VARIANT III - LIFE CYCLE COST BASIS
(IN THOUSANDS OF DOLLARS)

	FY89	FY90	FY91	FY92	FY93	FY94	FY95	FY96	FY97	FY98	TOTAL
NON-RECURRING COSTS											
Capture	5,564	1,614	0	0	0	0	0	0	0	0	7,178
Search/Image	1,102	9,058	3,579	2,809	1,964	1,839	0	0	0	0	20,350
SUBTOTAL	6,666	10,671	3,579	2,809	1,964	1,839	0	0	0	0	27,527
RECURRING COSTS											
Capture	394	11,641	19,438	18,883	17,629	15,629	8,878	8,094	7,161	6,165	113,911
Search/Image	0	0	5,725	6,337	6,824	7,274	7,438	7,275	7,082	6,824	54,779
SUBTOTAL	394	11,641	25,163	25,220	24,453	22,903	16,316	15,369	14,243	12,989	168,690
TOTAL COSTS	7,060	22,312	28,741	28,029	26,417	24,742	16,316	15,369	14,243	12,989	196,217
								Present value @		3.0%	168,579
										10.0%	122,555

Table 6. VARIANT IV - LIFE CYCLE COST BASIS
(IN THOUSANDS OF DOLLARS)

	FY89	FY90	FY91	FY92	FY93	FY94	FY95	FY96	FY97	FY98	TOTAL
NON-RECURRING COSTS											
Capture	5,430	1,614	0	0	0	0	0	0	0	0	7,044
Search/Image	1,347	12,673	3,519	2,072	2,072	1,839	0	0	0	0	23,522
SUBTOTAL	6,777	14,286	3,519	2,072	2,072	1,839	0	0	0	0	30,565
RECURRING COSTS											
Capture	394	11,641	19,438	18,883	17,629	15,629	8,878	8,094	7,161	6,165	113,911
Search/Image	0	0	5,780	6,302	6,646	7,023	7,193	7,085	6,956	6,784	53,770
SUBTOTAL	394	11,641	25,218	25,184	24,276	22,652	16,071	15,178	14,117	12,949	167,680
TOTAL COSTS	7,171	25,928	28,737	27,256	26,348	24,490	16,071	15,178	14,117	12,949	198,245
								Present value @		3.0%	170,659
										10.0%	124,645

Table 7. VARIANT V - LIFE CYCLE COST BASIS
(IN THOUSANDS OF DOLLARS)

	FY89	FY90	FY91	FY92	FY93	FY94	FY95	FY96	FY97	FY98	TOTAL
NON-RECURRING COSTS											
Capture	5,333	1,614	0	0	0	0	0	0	0	0	6,947
Search/Image	1,103	7,926	3,635	1,964	1,964	1,839	0	0	0	0	18,430
SUBTOTAL	6,436	9,540	3,635	1,964	1,964	1,839	0	0	0	0	25,376
RECURRING COSTS											
Capture	394	11,611	19,393	18,838	17,584	15,584	8,848	8,064	7,138	6,142	113,596
Search/Image	0	0	5,436	6,129	6,594	7,121	7,278	7,066	6,815	6,480	52,919
SUBTOTAL	394	11,611	24,829	24,967	24,178	22,704	16,126	15,130	13,954	12,622	166,515
TOTAL COSTS	6,830	21,151	28,463	26,931	26,142	24,543	16,126	15,130	13,954	12,622	191,891
									Present value @	3.0%	164,789
										10.0%	119,671

Table 8. VARIANT VI - LIFE CYCLE COST BASIS
(IN THOUSANDS OF DOLLARS)

	FY89	FY90	FY91	FY92	FY93	FY94	FY95	FY96	FY97	FY98	TOTAL
NON-RECURRING COSTS											
Capture	3,914	1,082	0	0	0	0	0	0	0	0	4,996
Search/Image	1,348	9,721	3,519	2,917	2,072	1,839	0	0	0	0	21,416
SUBTOTAL	5,262	10,803	3,519	2,917	2,072	1,839	0	0	0	0	26,412
RECURRING COSTS											
Capture	394	12,537	23,681	22,935	20,789	18,229	9,426	8,648	7,636	6,505	130,779
Search/Image	0	0	5,555	6,076	6,506	6,882	7,053	6,944	6,815	6,643	52,474
SUBTOTAL	394	12,537	29,236	29,011	27,295	25,111	16,478	15,592	14,451	13,149	183,254
TOTAL COSTS	5,656	23,340	32,755	31,929	29,367	26,950	16,478	15,592	14,451	13,149	209,666
							Present value @			3.0%	180,303
										10.0%	131,223

Table 9. VARIANT VII - LIFE CYCLE COST BASIS
(IN THOUSANDS OF DOLLARS)

	FY89	FY90	FY91	FY92	FY93	FY94	FY95	FY96	FY97	FY98	TOTAL
NON-RECURRING COSTS											
Capture	3,823	1,075	0	0	0	0	0	0	0	0	4,898
Search/Image	1,105	7,929	3,635	1,964	1,964	1,839	0	0	0	0	18,435
SUBTOTAL	4,928	9,004	3,635	1,964	1,964	1,839	0	0	0	0	23,332
RECURRING COSTS											
Capture	394	12,537	23,681	22,935	20,789	18,229	9,426	8,648	7,636	6,505	130,779
Search/Image	0	0	5,436	6,129	6,594	7,121	7,278	7,066	6,815	6,480	52,919
SUBTOTAL	394	12,537	29,116	29,064	27,383	25,349	16,704	15,714	14,451	12,986	183,699
TOTAL COSTS	5,322	21,541	32,751	31,028	29,346	27,188	16,704	15,714	14,451	12,986	207,031
								Present value @		3.0%	177,820
										10.0%	129,046

2) Increasing the rate at which the backlog is processed from 18,000 pages/day to 20,000 pages/day in the period FY91 to FY94 would result in the backlog being loaded one year earlier, at an increase in total cost of only about $1 million.

3) The LSS costs are primarily labor intensive. Using the figures for the Base Conceptual Design, the major costs contributors are:

Labor	70%
Hardware	16%
Facility	7%
Telecommunications ..	4%
Hardcopy production .	3%
Software	2%

4) Eliminating images on-line at the workstations (Variant III) is calculated to be more expensive than the base (with on-line images) due to the overriding expense of producing and shipping additional hardcopy documents. The Base Conceptual Design includes an assumption that approximately 10 million pages per year will be produced in hardcopy at the peak, and a 50% increase in hardcopy demand will be experienced if users do not have images on-line as in Variant III. On this basis, the additional hardcopy costs more than offset the reductions due to the elimination of on-line images. At no increase in hardcopy demand or half the level of hardcopy production, the opposite situation would exist; so the result is clearly dependent on the validity of the assumption (but not sensitive to it).

5) If images are required for on-line display as well as hardcopy production, it is less expensive to provide them on electronic (optical disk) format than on microfilm (Variant IV vs. Base). However, if on-line images are not required and a hardcopy production service is modeled after NTIS, the total cost is less than the base (Variant V vs. Base).

6) At a re-keying cost of $3.00 per page it is less expensive to set up an OCR production facility (Variant VI vs. Base).

7) Hardware full-text search (Variant II) is only slightly more expensive than the Base Conceptual Design (software full-text search). Given the ability of vendors to reduce hardware costs, this alternative could become less expensive in the future. However, the cost elements do not reflect higher program and schedule risk which would probably arise with this new technology. A detailed risk analysis should be undertaken if this variant is to receive further consideration.

5.0 BENEFITS ANALYSIS

The purpose of this benefits analysis is to provide a relative ranking of the LSS conceptual design variants based on the benefits they provide to the six usage groups identified in the Needs Analysis. The Base Conceptual Design will serve as the baseline system to which all system performance characteristics and user benefits will be compared. Cost differences between the Base Conceptual Design and the variants were not addressed as "benefits." Rather, differences in the costs of the systems were addressed in the cost analysis. Benefits derived from variations from the Base Conceptual Design which accrue to the user, and thus to the licensing process, are analyzed here. In all cases, the degree of support to licensing is considered the same, and the differences in effort (or avoidance of effort) to achieve that degree of support are compared.

A benefits sensitivity analysis is also presented in this section, which identifies user benefits that may be sensitive to system performance characteristics.

5.1 Methodology

The methodology employed in the system benefits analysis is a form of multi-attribute analysis known as the Relative Effectiveness Assessment Process (REAP). Individual alternative systems or concepts are compared with this process, and relative levels of benefit, or Measures of Effectiveness (MOEs), are determined for each system. The MOEs are based on the ability of the system to perform a particular set of functions that meet the stated requirements. Using this process, a relative ranking of systems is determined, using a set of weighted parametric evaluations.

A weighted parametric evaluation is performed by constructing a "tree" of parameters and evaluation criteria (Figure 3). At the base of the tree is the overall MOE or benefit derived from the system. This MOE is derived from a first level of parametric considerations, each with its own weight. The weights are used to combine the values for each consideration into the MOE; both arithmetic and geometric means can be used. In the same way, each primary consideration may be derived, in turn, from additional considerations. The "leaves" of the tree are the quantifiable variables which define and distinguish the systems. These are normalized to the values taken by a baseline system. As an example of the approach, if the system under consideration were an automobile, the MOE could be "overall driver satisfaction": the first level of parameters might include style, comfort, cargo space, and operations. "Style" might be a "leaf" and given a rating, "comfort" would be comprised of handling, ride, etc., while "operations" could be broken down into mean time between breakdown and mean cost to repair.

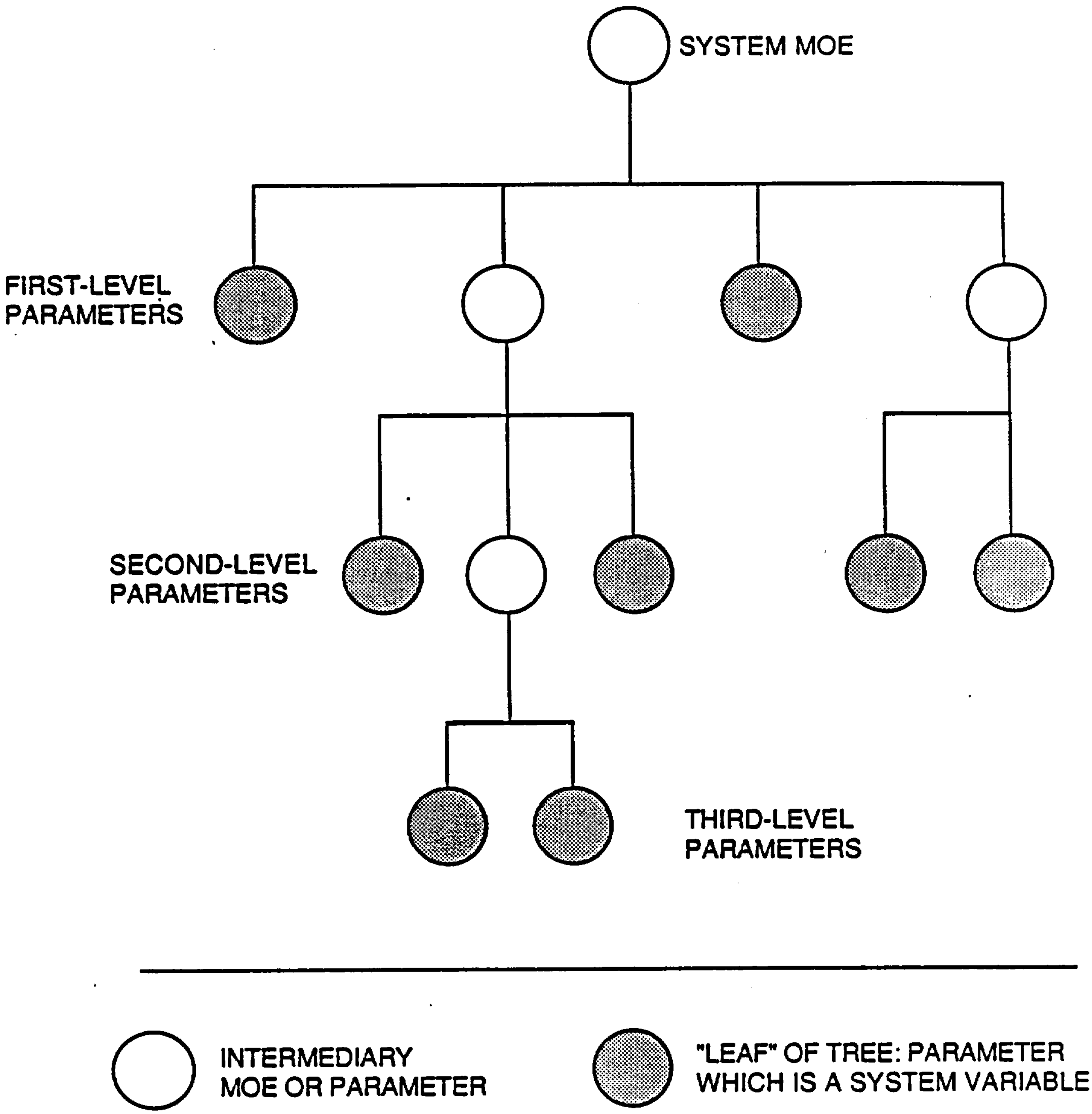


Figure 3. Tree Evaluation Parameters

This approach has several advantages over consideration of entire systems. First, it requires conscious selection and consideration of the relative importance of parameters. Secondly, by developing the hierarchical structure described above, each set of assignments can be made over a small, manageable decision set.

As the foregoing description suggests, the selection of the parameters to be used and the weights they are assigned determine the magnitude of the numerical outcome. Care has been taken both in selecting the parameters and weights, and in interpreting the results. Parameters were selected which distinguish among the alternatives, yet which do not conceal important similarities. In the automobile example, if the user is a general contractor, then the cargo space parameter may be of overriding importance. If, however, all the candidate vehicles have adequate storage, "cargo space" might be dropped from the list of criteria to accentuate the remaining differences to assist in selection. It would not be correct, however, to use these numerical values in a direct, unweighted comparison with cost.

In any information system, the primary bases of effectiveness are data retrieval and data quality (recall and precision): How does the user get information, how much is received, and how relevant is the question. In the case of the LSS, several aspects of the design approach are important. Each of the variants and the Base Conceptual Design have been engineered to provide equivalent data quality (in retrieval) and adequate retrieval speed. The variants differ primarily in their input methods, retrieval methods, and configuration. Accordingly, this analysis emphasized data retrieval. Any resulting requirements for additional equipment or personnel have been included on the cost side of the analysis.

The analysis tree derived for the LSS is shown in Figure 4. The LSS Base Conceptual Design was used for normalization. The overall benefit is composed of the benefit accruing to each of the six identified usage groups, as a weighted average. The weight assigned to each group is a combination of its size and importance (Section 5.2.1). The benefit seen by each group is the weighted average of eleven system function capabilities in searching, retrieval, printing, "capture delay", and image quality (Table 10). "Capture delay" refers to the amount of time between submittal of a document to the LSS and its on-line availability. The values of the parameters reflect the amount of time required to perform the functions described by the parameter a single time. Image quality was also expressed as time by treating it as the time required to determine if a printed image provides the required information.

The weights assigned to each of the eleven functional parameters differed from user group to user group. As the methodology requires, the weights remained constant across system variants. The values assigned to each of the parameters were determined by the expected characteristics of the system variant.

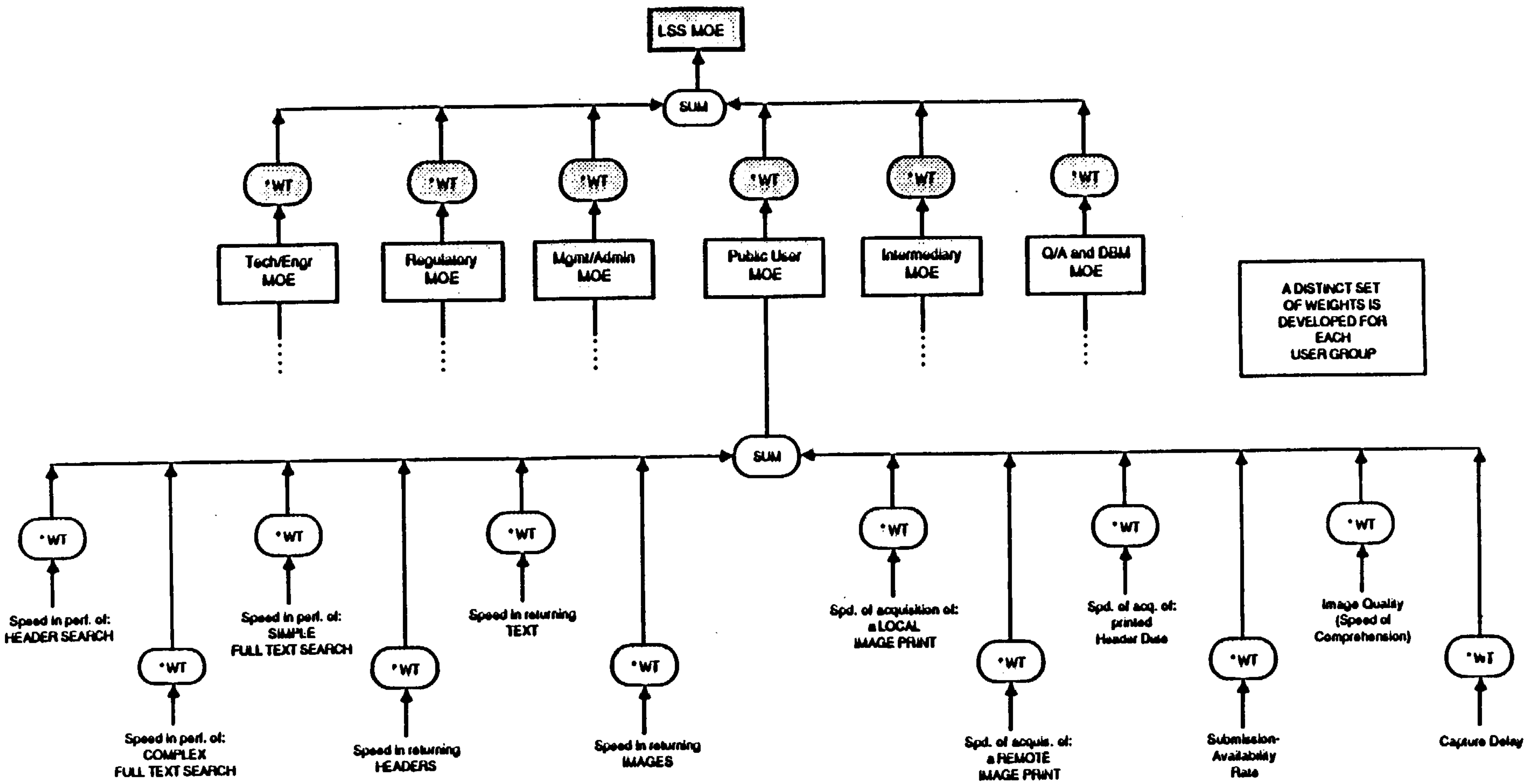


Figure 4. LSS Benefit Analysis Tree

TABLE 10. SYSTEM FUNCTIONAL PARAMETERS

Parameter Title	Description
Perform Header Search	Time to perform a search on document headers and return the number of "hits" and other statistics.
Perform Simple Full-Text Search	Time to perform a search on the complete text of the documents in the data base, using several logical combinations in the search definition, and return "hits" and other statistics.
Perform Complex Full-Text Search	Time to perform a search on the complete text of the documents in the data base, using a large number of logical combinations in the search definition, and return "hits" and other statistics.
Retrieve Header	Time to display a header for a document or a list of documents found by a search.
Retrieve Text	Time to display a page of text from a document or a list of documents found by a search.
Retrieve Image	Time to view an image found by a search. This is either the on-line display time at an image terminal in systems having such terminals, or the time to receive a remotely printed image.
Local Image Print	Time to print an image found by a search at an image printer local to the user. If local image printing is not supported in a variant, the value used is the time to receive a remotely-printed image.
Remote Image Print	Time to receive an image printed at a remote site and express mailed to the requestor.
Print Header Data	Time to receive (download) and print ASCII data from header files located by a search.
Capture Delay	Time which elapses between receipt of a document by the LSS data capture and loading system and its availability to users of the LSS.
Image Quality	Time required to scan a printed or displayed image for the purpose of determining if the quality of the image meets the user's needs.

5.2 Assignment of Weights and Values

The selection of parameters was followed by the equally important task of assigning values to those parameters and weighting the parameters to reflect their impact on overall system performance. Both the selection of parameters and the assignment of weights and values are interactive processes, which relied heavily on the participation of those familiar with the performance and requirements of the LSS system and those well versed in the application of the assessment methodology.

As implied in the automobile example in Section 5.1 and explained below in detail for the LSS, the assignment of weights and values influences the numerical outcome of the analysis. To ensure the outcome accurately models both system capabilities and user requirements, the weights and values assigned were derived from interviews with representatives of previously identified user groups and extensive interactions with groups of multi-disciplinary LSS experts. The subjects of the interviews were previously identified in the Preliminary Needs Analysis, and the multi-disciplinary experts include those involved in preparing the three LSS reports which precede this report. Thus the values and weights used in this analysis result from a consensus among well-informed individuals with appropriate backgrounds.

The employment of LSS experts in determining the weights assigned to LSS parameters selected for analysis provides a sound basis for conducting sensitivity analyses of these parameters. Sensitivity analysis provides yet another means of determining the influence of individual parameters on the overall benefits analysis and is thus a necessary part of this analysis.

Well established sensitivity analysis techniques were applied to the benefits analysis. The weights of the parameters were varied over a sufficiently wide range to determine the impact of individual parameters on the overall expression. This analysis indicated that major, reasonable changes in weight assignments within a user group varied the outcomes by approximately 5 percent.

5.2.1 Assignment of Weighting Factors

Each usage group will make different use of the LSS, and has different needs which the LSS must address. The overall benefit provided by the LSS is a weighted average of the benefit provided to each of the defined groups, as summarized in Table 11. Each group represents a fraction of the total user population, estimated in the Preliminary Needs Analysis. Each group also has a relative importance, or level of direct association with the licensing process, associated with its use of the LSS. The weight assigned to each user group in combining its MOE into the LSS MOE is a combination of these two factors.

TABLE 11. USAGE GROUP WEIGHTS

Usage Group	% of Total	Importance	Weight
Technical/ Engineering	45	1	45
Regulatory	25	3	75
Management/ Administration	5	2	10
Public	5	1	5
Intermediary	18	2	36
Q/A and Data Base Management	2	1	2

Within each user group, the weights assigned to each of the eleven system functions and parameters will also change. A short discussion of each function is provided below. The relative importance assigned to each function for each usage group is discussed in the subsequent sections. It must be emphasized that assigned weights have relevance only as relative factors <u>within</u> a user group: the weights need not be normalized across groups, and numerical comparison of normalized weights across user groups is not meaningful, since they are strictly relatable only to values at the same point on the tree.

The search processes (either on headers or full text) will return to the user the number of "hits" and other statistical information about the results of a query. Having concluded a series of searches, possibly using all search types, the user can then retrieve the headers associated with the selected documents, the text of the document (if any), or the document's image. The image may be accessible on-line or off-line depending on the system variant.

The documents may be printed from the image data either remotely (for large jobs) or locally (if the local system provides that capability). In addition, the ASCII data in the headers may be downloaded and printed locally.

A further weighing consideration is the amount of time that elapses between receipt of a document by the LSS and its availability for access by the system. This capture delay parameter will probably be of greater importance to the administrative and regulatory users than to the technical personnel.

Finally, image quality is of some importance. However, since the minimum acceptable quality of a stored document image will be very high by most standards, the overall value of document quality importance is low: it is not expected to be an issue with most users.

5.2.1.1 Technical/Engineering Usage Group

These users are interested in retrieving primary data, published analyses of technical issues, and descriptions of procedures and methods. Since much of this information is graphic in nature, technical and engineering users will place a premium on image retrieval, especially in the context of header searches (which must be used for graphic intensive material). As technically-oriented people, they will make the greatest use of the header information, but will utilize the full text search capabilities when their primary goal is textual. The print requirements of the technical user are primarily associated with remote printing of full documents and of local header data printing. Engineering usage of the database is expected to be primarily archival, hence the low relative importance assigned to minimizing capture delay. The relative weights assigned to each of the functions are shown in Table 12.

TABLE 12. WEIGHTS ASSIGNED TO FUNCTIONS FOR TECHNICAL/ENGINEERING USAGE GROUP

Function	Weight
Perform Header Search	12
Perform Simple Full-Text Search	6
Perform Complex Full-Text Search	3
Retrieve Header	10
Retrieve Text	4
Retrieve Image	9
Local Image Print	2
Remote Image Print	8
Print Header Data	8
Capture Delay	0
Image Quality	3

5.2.1.2 Regulatory and Licensing Support Usage Group

The focus of these lawyers and licensing engineers on textual material, in the form of reports, minutes of meetings, correspondence, and regulations and regulatory guidance indicates that they will perform most of their searches <u>via</u> the full-text search capability of the LSS. They will be primarily interested in using this search method to pull out text, and, to a lesser extent, images. Their use of header searches is expected to be concentrated on identifying text data, but they will also be interested in header information in the form of "from-to data". Their print requirements are similar to the technical users, but this user group has a relatively higher need for local image printing. Maintaining an up-to-date data base is highly important and highly visible to these users. The relative weight assigned to each of the system functions is shown in Table 13.

TABLE 13. WEIGHTS ASSIGNED TO FUNCTIONS FOR REGULATORY USAGE GROUP

Function	Weight
Perform Header Search	5
Perform Simple Full-Text Search	12
Perform Complex Full-Text Search	2
Retrieve Header	2
Retrieve Text	10
Retrieve Image	6
Local Image Print	5
Remote Image Print	10
Print Header Data	2
Capture Delay	10
Image Quality	2

5.2.1.3 Management/Administrative Usage Group

The background, approach, and interest of these users is quite similar in the regulators discussed above, but their needs for search types are less skewed towards text. They are expected to make greater use of planning documents and charts. The average document accessed will be small, and the need for hard copy will be high, giving a strong weight to local image printing. In general they will tend to know the identity of the documents they need, lowering the importance attached to searching. The relative weight assigned to each of the system functions is shown in Table 14.

TABLE 14. WEIGHTS ASSIGNED TO FUNCTIONS FOR MANAGEMENT/ADMINISTRATION USAGE GROUP

Function	Weight
Perform Header Search	4
Perform Simple Full-Text Search	6
Perform Complex Full-Text Search	1
Retrieve Header	1
Retrieve Text	5
Retrieve Image	3
Local Image Print	10
Remote Image Print	5
Print Header Data	4
Capture Delay	6
Image Quality	4

5.2.1.4 Public Information and General Public Usage Group

This group will be using the LSS in support of the information needs of the general public, and they will be dealing primarily with descriptive information, OCRWM activities, and summary data on technical and environmental issues. They are expected to use primarily full-text searching, with major interest in retrieving text as opposed to headers or images. Most members of this group will have less frequent or less easy access to the search capabilities, which increases the importance of their printing requirements. Since many of these users will be following and monitoring the regulatory and licensing processes, they will have a strong interest in an up-to-date data base. The relative weight assigned to each of the system functions is shown in Table 15.

TABLE 15. WEIGHTS ASSIGNED TO FUNCTIONS FOR PUBLIC USAGE GROUP

Function	Weight
Perform Header Search	2
Perform Simple Full-Text Search	10
Perform Complex Full-Text Search	2
Retrieve Header	3
Retrieve Text	9
Retrieve Image	5
Local Image Print	8
Remote Image Print	10
Print Header Data	4
Capture Delay	8
Image Quality	4

5.2.1.5 Intermediaries Usage Group

The intermediary users perform search and retrieval services on a professional basis for members of other usage groups. The intermediary can be characterized by highly professional use of searching techniques, both of headers and full text, with concomitant major use of statistics and header retrieval. As these users are not themselves the final consumers of the documents located, they have relatively low requirements for text and image retrieval, with correspondingly high needs for rapid printing. The relative weight assigned to each of the system functions is shown in Table 16.

TABLE 16. WEIGHTS ASSIGNED TO FUNCTIONS FOR INTERMEDIARY USAGE GROUP

Function	Weight
Perform Header Search	12
Perform Simple Full-Text Search	10
Perform Complex Full-Text Search	8
Retrieve Header	10
Retrieve Text	3
Retrieve Image	3
Local Image Print	10
Remote Image Print	12
Print Header Data	8
Capture Delay	4
Image Quality	3

5.2.1.6 Quality Assurance (Q/A) and Database Management Usage Group

These users work directly with the LSS database and its support systems. As users who will make use of both the LSS searching system and special systems software to perform their jobs, their use of the LSS will be primarily statistical in nature, with the contents of documents being of relatively little interest. By the same token, their printing requirements, with the exception of header data, are also low. The relative weight assigned to each of the system functions is shown in Table 17.

TABLE 17. WEIGHTS ASSIGNED TO FUNCTIONS FOR Q/A AND DATABASE MANAGEMENT USAGE GROUP

Function	Weight
Perform Header Search	4
Perform Simple Full-Text Search	4
Perform Complex Full-Text Search	2
Retrieve Header	1
Retrieve Text	1
Retrieve Image	1
Local Image Print	1
Remote Image Print	2
Print Header Data	2
Capture Delay	2
Image Quality	1

5.2.2 Assignment of Parameter Values and Calculation of Benefit

The input parameters were selected in the benefits analysis in a way which differentiates between the variants, based on functions which were considered important to the six usage groups. These eleven parameters are listed in Table 10 and the variables and MOE formulas are summarized in Figure 5. These parameters all represent average times to perform certain functions, such as performing searches or receiving data. Shorter time values represent greater benefit to the user, therefore these parameters are converted to "speed" values (i.e., one over time).

By converting input time parameters to speed, the higher values indicate better performance and therefore greater user satisfaction. The result is a list of eleven parameters which define the average speed at which a system variant is expected to perform those functions. For the Base Conceptual Design and each of the seven variants, the speed values were first determined for most of the eleven input parameters using $S_i^k = 1 / T_i^k$, where T_i^k represents the i^{th} input function for the k^{th} variant, and S_i^k represents the resulting speed. Two variables, Retrieve Image and Local Image Print, had very wide ranges of values due to the presence of variants both with and without on-line image retrieval. The speed value for these two variables was determined using a piecewise continuous exponential function which more accurately reflects user perceptions of "effective" speed in receiving information: $S_i^k = H(T_i^k)$. The speeds are then normalized to the Base Conceptual Design in order to obtain relative values. This is done by dividing the value of each S_i^k by S_i^B, the corresponding speed of the parameter in the Base Conceptual Design. This calculation results in relative speeds where a value of 1.0 represents the speed of the function in the Base Conceptual Design, values less than 1.0 are slower and provide less user satisfaction than the Base Conceptual Design, and values greater than 1.0 are faster and provide greater user satisfaction than the Base Conceptual Design.

The normalized values are used in combination with the weighting factors to obtain Measures of Effectiveness (MOEs) for each user group in each variant by using the weighted arithmetic mean of the eleven normalized parameters: each normalized parameter is multiplied by the relative weight associated with the i^{th} parameter of the j^{th} user group. The sum of these products are divided by the sum of the weights to yield the MOE or relative benefit of the variant, MOE_j^k (Figure 5). The Base Conceptual Design always receives a relative MOE of 1.0.

Finally, the relative benefit of each variant is calculated by taking the weighted arithmetic mean of the MOE of the usage groups for that variant. Weights associated with each usage group, W_j, are multiplied by each usage group MOE_j^k and the sum is divided by the sum of the user group weights. The relative benefit (MOE) of each variant is then compared with the Base Conceptual Design, which has an MOE of 1.0. Values greater than 1.0 indicate that the variant provides greater overall benefit to users than the Base Conceptual Design, and values less than 1.0 indicate less overall benefit. These MOE values can be used to rank the variants according to their ability to satisfy user requirements based on the set of input parameters used to differentiate the variants.

Sensitivity analyses were then performed on the MOEs by varying the values of the input parameters within a reasonable range, since they represent estimates of the performance of the conceptual systems. The results of these analyses are described in the following sections, in the form of a "best estimate" used as a central value for the relative benefit and low and high values based on excursions about the central value. A short description of each variant is included for reference. Complete descriptions of the features of each system can be found in the Conceptual Design Analysis.

FIGURE 5. BENEFIT ANALYSIS INPUT PARAMETERS AND FORMULAS

VARIABLES
(See Table 10 for detailed descriptions of each input parameter.)

T_{hds} = The average time to perform a header search.
T_{sft} = The average time to perform a simple full-text search.
T_{cft} = The average time to perform a complex full-text search.
T_{rhd} = The average time to retrieve and display header data.
T_{rtx} = The average time to retrieve and display text.
T_{rim} = The average time to retrieve and display (or deliver) images.
T_{lip} = The average time to receive a local image print.
T_{rip} = The average time to receive a remote image print.
T_{hdp} = The average time to receive a header data printout.
T_{cap} = The average document capture delay time.
T_{qim} = Quality of image expressed as the average time to review/accept a page.

MEASURES OF EFFECTIVENESS FORMULAS

For i = 1..5, 7..11, $S_i^k = 1 / T_i^k$; $S_{rim} = H(T_{rim})$; $S_{lip} = H(T_{lip}).5$

(see explanation in text)

$$MOE_j^k = \frac{SUM(i=1...11)[S_i^k * W_{ij} / S_i^B]}{SUM(1=1...11)W_{ij}}$$

$$MOE^k = \frac{SUM(j=1...6)[MOE_j^k * W_j]}{SUM(j=1...6)W_j}$$

FIGURE 5. BENEFITS ANALYSIS INPUT PARAMETERS AND FORMULAS (Continued)

Where

S_i^k is the speed of the i^{th} system function for the k^{th} variant as defined above.

MOE_j^k is the MOE of the j^{th} user group of the k^{th} variant.

MOE^k is the MOE of the k^{th} variant.

W_{ij} is the weight of the i^{th} function for the j^{th} user group for calculating the MOEs for user groups within variants.

W_j is the weight of the j^{th} user group used for calculating the MOE for a variant relative to the baseline conceptual design.

i indexes a system function:

i = 1...11 =	hds, sft, cft, rhd, rtx, rim, lip, rip, hdp, cap, qim	(subscripts tied to parameters at the beginning of this figure)

j indexes a user group:

j = 1...6 =
- T : Technical/Engineering
- R : Regulatory/Licensing Support
- M : Management/Administrative
- P : Public
- I : Intermediaries
- Q : Q/A and Database Management

k indexes a system variant:

k = 1...8 =
- B : Base Conceptual Design
- I : Variant I
- II : Variant II
- III : Variant III
- IV : Variant IV
- V : Variant V
- VI : Variant VI
- VII : Variant VII

5.2.2.1 Base Conceptual Design

The Base Conceptual Design for the LSS is a single node system which provides on-line access to the information required to support the licensing process to construct and operate a nuclear waste geologic repository as specified by the negotiated rulemaking process. The access consists of on-line search capability of bibliographic fields (referred to as headers) and full ASCII text of each document in the system. The headers, text, and images can be viewed on-line via special workstations. The Base Conceptual Design includes the hardware, software and operations necessary to capture the information, load it into the system, store it, provide on-line query and display, and to distribute hardcopy upon request. The LSS is comprised of the capture system, search system, image system, communications and workstations. A more complete description of the Base Conceptual Design is provided in Section 3.1 of this report.

The Base Conceptual Design has, by definition, a relative benefit value of 1.0.

5.2.2.2 Variant I - Full Replicated Nodes

Variant I differs from the Base Conceptual Design in that it has two fully replicated search and image system nodes, one located in Washington, DC and one in Las Vegas, NV. (Section 3.2.1 contains a more detailed description of Variant I.) Each node uses a machine sized between 50% and 100% of that of the Base Conceptual Design. The overall "horsepower" available to the users is therefore greater than or equal to that of the Base Conceptual Design, so that all parameters which are affected by machine size are affected. These include the three "time to search" and "time to retrieve" parameters, as well as the retrieve portions of the time to print local images and header data. These distinctions can be summarized as a "load factor" relative to the Base Conceptual Design, which measures the "headroom" each computer has: If two Base Conceptual Design computers are purchased then the load factor is 0.5; if two computers with half the capacity are purchased the load factor is 1.0. The other parameter changed in Variant I is the Capture Delay, which now must deal with the extra configuration management requirements associated with two databases.

The best estimate for the load factor was based on computer modeling performed as part of the design analysis. Each of the two computers will be slightly "oversized", giving a load factor of 0.8. Excursions were set to the limits discussed above. Capture Delay was increased by 20%, with excursions at 7% and 33%.

The results of the sensitivity analyses are shown in Figure 6. Relative benefit ranges from 0.98 to 1.48, with load factor being the driver. The central value is 1.11. This system shows the greatest range of impact on benefit, and it is the only variant with a benefit greater than 1.0.

System Variant / Identifying Characteristics	Benefit: Low	Benefit: Central	Benefit: High
Base Conceptual Design		1.00	
Variant I: Two Nodes			
Load Factor	0.98	1.11	1.48
Capture Delay	1.10	1.11	1.12
Variant II: Hardware Full Text Searching			
Full Text Search Times	0.88	0.90	0.98
Capture Delay	0.87	0.90	0.93
Variant III: No Level 2 Workstations			
Image return time	0.74	0.85	1.00
Variant IV: On-line digitizing of Microform			
Image return time	0.95	0.95	0.95
Capture Delay	0.94	0.95	0.96
Image Quality	0.94	0.95	0.96

Graphically

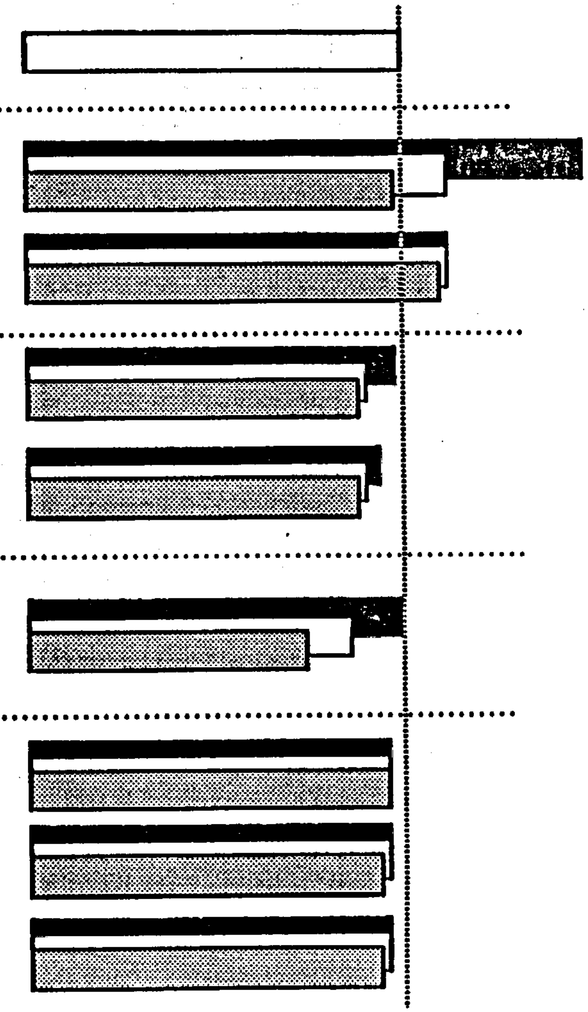


Figure 6. LSS BENEFITS SUMMARY AND SENSITIVITY ANALYSIS

System Variant Identifying Characteristics	Benefit: Low	Benefit: Central	Benefit: High
Base Conceptual Design		1.00	
Variant V: Microform off-line			
Image return time	0.70	0.76	0.96
Capture Delay	0.76	0.76	0.77
Image Quality	0.75	0.76	0.77
Variant VI: Text Re-Keyed			
Capture Delay	0.95	0.96	0.99
Variant VII: Comination III, V, VI			
Image return time	0.69	0.75	0.95
Capture Delay	0.74	0.75	0.77
Image Quality	0.74	0.75	0.76

Graphically

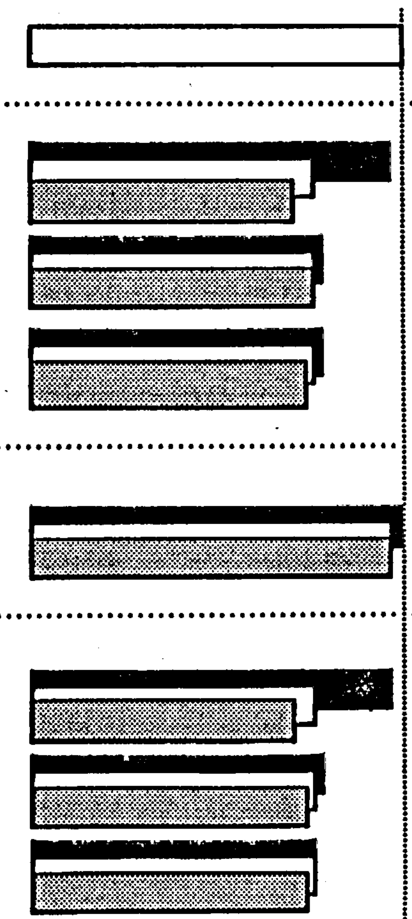


Figure 6. LSS BENEFITS SUMMARY AND SENSITIVITY ANALYSIS (Continued)

5.2.2.3 Variant II - Hardware Full-Text Search

In the Base Conceptual Design full-text search is implemented <u>via</u> storage of the full text plus creation of an inverted index. The full-text data base software uses both to respond to user queries. Variant II replaces them and the text data base manager with specialized hardware processors. All other functions (E-mail, header searches, <u>etc</u>.) performed by the search system host computer remain the same as the Base Conceptual Design. Section 3.2.2 contains a more detailed description of Variant II.

The parameters affected are the full-text search times (both complex and simple searches take the same time) and Capture Delay, which may be slightly improved since no indexes need be built on the text.

The central values and their excursions for the hardware full-text search are based on assessments of existing and near-term technology performed in the design analysis. A fairly broad range of values was used in the excursion to reflect uncertainties in this area.

The relative benefit, as shown in Figure 6, ranges from 0.87 to 0.98, with the central value at 0.90. As in all variants, changes in Capture Delay had minimal impact on relative benefit.

5.2.2.4 Variant III - Images Are Not Supported At Workstations

In this variant from the Base Conceptual Design the capability to view electronic (bit-mapped) images on the screen at the Level 2 workstations is excluded. Images are still stored on-line using optical disks at the LSS, but print jobs must be sent to users after they are made at the LSS site. To compensate for not being able to display images at a workstation, the requests for printing images would increase.

This variant affects the Local Image Print and Retrieve Image parameters, which must be set at the value used for Remote Image Print. A central value of two days with excursions to one and seven days were used, giving relative benefits from 0.84 through 1.00. The relative benefit of 1.00 was achieved by reducing image return time to one day. This is an improvement over the Base Conceptual Design image return time of two days, and is achieved by methods not directly or entirely related to data processing capability, but rather to the time required to physically transport a hardcopy image.

5.2.2.5 Variant IV - Microform Digitizers in Capture and Image Systems

In this variant on the Base Conceptual Design the changes occur in the capture and image systems. Microform digitizers are used to create the OCR input for documents available only on microform. Microform replaces optical disks for the storage and retrieval of images. Hardcopy images are still scanned to form bit-mapped images which are OCRed to create text files for searching, but microform images are created from hardcopy for image storage.

User retrieval of stored images requires the accession of microform images from storage cassettes and on-line digitization before transmission to image workstations.

Three parameters are affected in this variant: Image Return Time, Capture Delay, and Image Quality, all associated with the use of on-line microform. Image return is slowed due to the need to load a reel and fast-forward to the requested image; Capture Delay is increased due to the need to develop and work with microform; and Image Quality is slightly degraded due to long-term stretching of the film. Variations about the central value for all of these parameters had minimal effect on the relative benefit, which had a central value of 0.95.

5.2.2.6 Variant V - Microform Off-Line Image Storage and Retrieval

Variant V of the Base Conceptual Design replaces the on-line image system with an off-line service for obtaining hardcopy or microform copies of LSS documents. All image storage and retrieval is performed off-line using microform. The capability to view electronic (bit-mapped) images on the screen at Level 2 workstations is excluded, as in Variant III. The affected parameters are the same as those of Variant IV.

The central values selected give a relative benefit of 0.76, which is primarily driven by the Image Return time, as shown in Figure 6. Excursions range from 0.70 to 0.96, with the highest value resulting from a one-day turnaround in printing time: e.g., a request submitted before noon would be sent in the same day's overnight delivery service. It should be noted that this one day image return time is a full day less than the image return time designated for the Base Conceptual Design.

5.2.2.7 Variant VI - Full Text *via* Re-keying

In this variant there is no automated text conversion (OCR) process. The conversion of hardcopy text to ASCII is accomplished by re-keying the document. An expected 99.8% accuracy of data *via* re-keying would be achieved by double keying the original source document. Since the text conversion will be accomplished *via* re-keying there will be no requirement for optical character recognition equipment and associated software. However, the re-keyed documents would require processing through a digital scanning device since bit-mapped image capture and storage is required. This affects only the capture delay of documents, as image storage and retrieval remain unchanged.

Capture Delay was increased by a factor of 2, for the central value, based on a need to ship documents to areas where low-cost keyboard entry was available. Excursions from a range of just over the Base Conceptual Design value to a factor of 2.7 were also used in the analysis of this variant. The relative benefit of 0.96 and the excursions to 0.95 and 0.99 are shown in Figure 6.

5.2.2.8 Variant VII - Combined Variants III, V and VI

Variant VII combines hardware and software changes for the Level 2 workstations (Variant III), the removal of the on-line image system (Variant V) and the re-keying of all documents instead of OCR (Variant VI). The affected variables are Image Return Time (Variants III and V), Capture Delay (Variants V and VI) and Image Quality (Variant V). The ranges used in each case were a combination of the values used for the variants combined.

As seen in Figure 6, this variant is very similar to Variant V, as the use of off-line storage and retrieval is the primary driver. The central value for the relative benefit of 0.75 is, with Variant V, the lowest found.

5.3 Benefits Analysis Summary

As discussed in Section 5.1, the numerical values must be considered in light of the parameters selected. The parameters used in this analysis were selected with the intent of highlighting the benefits associated with distinctions between the Base Conceptual Design and variants. Overall system performance, as measured by the search criteria, is constant across most of the systems, and serves to keep the distinctions in perspective. Some general conclusions from the analysis are:

1) The single biggest factor in increasing user benefits is the lowered load factor in Variant I: the more computer power available to each user, the better. (To a certain extent this is true whether the computers are operating as single or double nodes.) Note that if the total computing capacity is kept equal to the Base Conceptual Design (i.e., the low benefit load factor sensitivity for Variant I), then benefit decreases to somewhat less than base.

2) The use of off-line image storage and retrieval significantly reduces relative benefit unless the average time to receive a print request is reduced to one day; in that case, the impact is small (cf. the high benefit image return time sensitivity for Variant V).

3) Over the ranges examined, Capture Delay and Image Quality have only minimal impact on relative benefit in this analysis. Re-keying text (Variants VI and VII) is comparably useful to the Base Conceptual Design, in spite of the longer Capture Delay. Off-line microform storage of images (Variants V and VII) is appreciably less useful because of Image Return Time, not Image Quality.

6.0 COMPARATIVE COSTS/BENEFITS

Table 18 below summarizes, in relative terms, the results of the cost and benefit analyses among the alternative configurations.

TABLE 18. COST AND BENEFIT SUMMARY

Alternative	Relative Cost	Relative Benefit
Base	1.00	1.00
Variant I	1.21	1.11
Variant II	1.01	0.90
Variant III	1.01	0.85
Variant IV	1.02	0.95
Variant V	0.98	0.76
Variant VI	1.08	0.96
Variant VII	1.06	0.75

Considering the results of both the cost and effectiveness analyses, it would appear that the Base Conceptual Design offers the best combination of cost and performance, while also, as noted in the Conceptual Design Analysis, offering the lowest overall risk to the program. Variants IV and VI have associated costs/benefits comparable to that of the Base Conceptual Design, within the probable uncertainties of this analysis under the conditions examined.

From the standpoint of costs, only one alternative, Variant V (no on-line images, hardcopy from microfilm), offered a lower total life-cycle cost than the base. The difference is actually not significant (being only 1.6%) and is within the error bounds of the calculations themselves. Nevertheless, when taken with a perceived measure of effectiveness of only 76% of the base, the possible slight cost savings of Variant V does not seem to warrant the loss in performance relative to the base. It should also be noted however, that in the sensitivity study associated with the benefits calculations, Variant V was perceived to be almost (0.96) as effective as the base if the image (hardcopy) could be provided in 1 day or less. This would indicate that a more detailed review of this alternative may be warranted to determine if a more cost effective design could be produced on this concept, while still providing effective results.

From the benefit side, only one alternative, Variant I, appeared to exhibit an improvement in effectiveness over the base. This improvement was not inherent in the fact that the variant was based on two operating systems, but from the fact that the two systems combined provided more computing capacity per user than the base (i.e., 1.6 times the capacity). At equal capacity, Variant I has slightly less benefit than the base. One could conclude, therefore that improved system effectiveness could be obtained merely by increasing computing capacity, without the need for the two redundant data bases included in Variant I. This would suggest that while it may not appear cost effective to consider Variant I, since it produces only an 11% improvement in performance at a cost increase of 21%, the LSS design should include an analysis of computer performance versus perceived benefits in order to optimize the search system computer size determination.

As noted in Section 4.2, the cost analysis demonstrates that not having on-line images (Variant III) is more expensive than the Base Conceptual Design. This is due to the cost of providing additional hardcopy pages to the users (which is to be expected when on-line images are not available). The cost of additional hardcopy production is a function of: 1) the level of hardcopy production assumed; 2) the percent increase in hardcopy demand without on-line images; and 3) the cost per page to provide the hardcopy. Since these are all assumptions in the calculations and are subject to uncertainty, it is possible to postulate conditions under which it might become less expensive not to have on-line images (for example if there is no increase in hardcopy demand). However, unless it is assumed that demand for hardcopy actually decreases without access to on-line images, in the limit Variant III could only be $2 million less that the base which is the cost of providing the on-line images. This difference is only a 1 percent of the total life cycle costs. Considering the benefit side of the picture, this 1 percent reduction is associated with a relative measure of effectiveness of 0.85 for Variant III. Therefore even if the hardcopy figures are not accurate, it does not appear to be cost effective to eliminate on-line images considering that the 1 percent potential savings corresponds to a 15 percent decrease in system effectiveness.

7.0 SENSITIVITY ANALYSIS

The costs for the Base Conceptual Design and the seven variants were computed based on the best estimates of the: 1) LSS data volume (DOE, 1988c); 2) number of expected simultaneous on-line users; and 3) percentage of text to be available for full text searching. These best estimates were derived from the Needs and Data Scope analyses (DOE 1988a, DOE 1988b). To determine if the relative costs of the designs depended on any of these three "volume" variables, a sensitivity analysis was conducted. Six sensitivity cases were created by developing two additional estimates for each of the three variables. Costs were computed for each case for each variant, a total of forty-eight individual variations.

7.1 Methodology

For each of the volume variables, two additional estimates were derived using range data from the Needs and Data Scope analyses. Sensitivity cases were created by changing one variable at a time. This resulted in six independent sensitivity cases. For each case and each variant, operations were reviewed to determine the number of staff and amount of equipment necessary to support the workload resulting from the modified volume estimate. Search, image and communications systems were re-sized and re-costed. The resulting staffing levels, equipment costs and processing loads were input to the cost model described in Section 4 to compute the total life-cycle costs associated with each combination of sensitivity case and variant. The resulting costs are presented in Section 7.2.

The following three digit numbering scheme is used to identify the six sensitivity cases:

> The left most digit identifies which data volume estimate is used (Section 7.1.1).
>
> The middle digit indicates the percentage of text to be available for full-text search (Section 7.1.2).
>
> The right most digit indicates the number of simultaneous users (Section 7.1.3).

In the following description, the number 1 indicates the nominal value is used. The numbers 2 and 3 identify the two deviations from the nominal. For example, case 111 is comprised of the nominal values for all three variables, 211 is comprised of the first deviation for data volume and nominal values for percent of full text to capture and the number of simultaneous users.

7.1.1 Volume of Data

The nominal value used for the data to be loaded was the high estimate of the number of pages to be processed by the end of the ten year analysis period (1998) as determined by the Preliminary Data Scope Analysis (DOE, 1988b) and updated in the Appendix B of the Conceptual Design Analysis (DOE, 1988c) report. The volume of data determines four parameters used in the cost model: the number of documents to be processed, the number of pages to be processed and the amount of magnetic and optical storage space required.

The first non-nominal data volume value is the low estimate of the number of pages from the Appendix B of the Conceptual Design Analysis based on the Preliminary Data Scope report. The low estimate also has different factors for the average number of pages per document and the average number of characters on a page.

The second non-nominal data volume value is 50 percent of the high estimate of the number of pages. The pages per document and characters per page factors are the same as for the analysis with the high estimate.

Case	Description	No. of pages Thru 1998	Total ASCII Mbytes (*)	Images Gbytes
111	Data volume high	27,921,000	200,194	1,396
211	Data volume low	21,404,000	80,907	749
311	Data volume 1/2 high	13,960,500	100,097	698

* The ASCII storage for headers, ASCII text and both indices

7.1.2 Percentage in Full-Text Search

The nominal value for the percentage of text to be processed and available for on-line searching is 100 percent of the materials that are ASCII text. The two non-nominal values used are 50 percent and 25 percent of the ASCII text to be processed and made available for on-line searching. All the percentages are based on the amount of ASCII text associated with the data volume high estimate.

Case	Description	% of ASCII Text	Total ASCII Mbytes (*)
111	100% Full Text	100	200,194
121	50% Full Text	50	112,662
131	25% Full Text	25	68,896

* The ASCII storage for headers, ASCII text and both indices

7.1.3 Number of Users

This variable is the maximum number of on-line simultaneous users at peak system loading. It is a key variable in determining the amount of computer capacity the search system must provide. The communications loads, number of Level 2 workstations, and amount of hardcopy to be requested are also affected. The nominal number is 100 simultaneous users. An upper value of 175 users was selected to represent the maximum based on the estimated number of users during a peak load period in the Preliminary Needs Analysis report. Half of the nominal number was selected to represent a lower limit.

Case	Maximum no. of simultaneous users
111	100
112	175
113	50

7.2 Results

The total life-cycle costs for each sensitivity case, design variant combination is shown in Table 19. Table 20 shows the percentage difference of each variant-sensitivity case compared to the equivalent base-sensitivity case. The value at the bottom of each column in that table is the average total life-cycle cost differences for the variants from the total life-cycle cost of the base design. An examination of the table indicates that all variants, except for Variant I (which has two search/image sites), are within -3.4 to +7.5 percent of the base design's cost. Only Variant V's cost is consistently less than the base, but only by an average of 2.2 percent. The two-site variant (I) is on the average 20.2 percent more expensive. Within each variant the largest range of percentage differences from the base is 6.8.

Table 21 shows the percentage difference for each sensitivity case within a variant from the nominal case for the variant. The figures show that each sensitivity has nearly the same percentage change for each variant.

Analysis of the results presented in Table 19 and their associated cost model intermediate results leads to the following observations.

1) Overall, none of the sensitivity cases show any significant change from the results obtained based on the nominal case discussed in Sections 4 and 5. Therefore, the conclusions reached in Section 6 remain valid.

2) Over the ranges examined, the volume of information to be processed has the largest effect on total life-cycle costs.

TABLE 19. RESULTS OF LIFE CYCLE COST SENSITIVITY STUDY
(000's ommitted)

SENSITIVITY		BASE	VAR I 2 sites	VAR II Hdwe FT	VAR III No image	VAR IV Micro	VAR V NTIS	VAR VI Rekey	VAR VII
111		194,966	235,686	197,288	196,217	198,245	191,891	209,666	207,031
211	vol low	150,119	181,596	154,135	150,382	152,189	146,392	159,660	156,287
311	vol 1/2 hi	129,827	158,592	132,960	129,481	132,122	125,452	136,767	132,641
121	50% F T	169,124	205,206	172,071	170,379	170,810	166,336	175,328	172,977
131	25% F T	156,705	186,823	160,055	157,956	156,791	154,380	157,757	155,871
112	175 sim usr	215,867	249,787	211,226	214,053	219,146	211,447	228,965	226,587
113	50 sim usr	185,853	224,877	188,698	187,100	189,132	181,141	199,504	196,280

TABLE 20. PERCENTAGE VARIANCE FROM THE BASE CONCEPTUAL DESIGN

SENSITIVITY		BASE	VAR I 2 sites	VAR II Hdwe FT	VAR III No image	VAR IV Micro	VAR V NTIS	VAR VI Rekey	VAR VII
111		0.0%	20.9%	1.2%	0.6%	1.7%	-1.6%	7.5%	6.2%
211	vol low	0.0%	21.0%	2.7%	0.2%	1.4%	-2.5%	6.4%	4.1%
311	vol 1/2 hi	0.0%	22.2%	2.4%	-0.3%	1.8%	-3.4%	5.3%	2.2%
121	50% F T	0.0%	21.3%	1.7%	0.7%	1.0%	-1.6%	3.7%	2.3%
131	25% F T	0.0%	19.2%	2.1%	0.8%	0.1%	-1.5%	0.7%	-0.5%
112	175 sim usr	0.0%	15.7%	-2.1%	-0.8%	1.5%	-2.0%	6.1%	5.0%
113	50 sim usr	0.0%	21.0%	1.5%	0.7%	1.8%	-2.5%	7.3%	5.6%
	Average		20.2%	1.4%	0.3%	1.3%	-2.2%	5.3%	3.5%

TABLE 21. PERCENTAGE VARIANCE FROM NOMINAL CASE

SENSITIVITY		BASE	VAR I 2 sites	VAR II Hdwe FT	VAR III No image	VAR IV Micro	VAR V NTIS	VAR VI Rekey	VAR VII	Average
111		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
211	vol low	-23.0%	-23.0%	-21.9%	-23.4%	-23.2%	-23.7%	-23.9%	-24.5%	-23.3%
311	vol 1/2 hi	-33.4%	-32.7%	-32.6%	-34.0%	-33.4%	-34.6%	-34.8%	-35.9%	-33.9%
121	50% F T	-13.3%	-12.9%	-12.8%	-13.2%	-13.8%	-13.3%	-16.4%	-16.4%	-14.0%
131	25% F T	-19.6%	-20.7%	-18.9%	-19.5%	-20.9%	-19.5%	-24.8%	-24.7%	-21.1%
112	175 sim usr	10.7%	6.0%	7.1%	9.1%	10.5%	10.2%	9.2%	9.4%	9.0%
113	50 sim usr	-4.7%	-4.6%	-4.4%	-4.6%	-4.6%	-5.6%	-4.8%	-5.2%	-4.8%

3) Change in total data volume processed has more cost impact than a corresponding change in the percentage of full text to be available. This is because the OCR process and equipment costs are a small portion of the total capture system costs.

4) As the amount of text to be processed is reduced, the investment in OCR equipment becomes less cost effective, and the OCR process becomes less of an advantage over re-keying.

8.0 CONCLUSIONS

The 10 year life-cycle costs for a number of alternative LSS designs which meet the needs and requirements of the users and the Negotiated Rulemaking Process have been evaluated. With a few exceptions, the costs are very similar because the predominant cost (data capture operations) is not greatly affected by the alternative designs and because the designs themselves cannot vary dramatically and still meet the stated requirements.

The costs, on the order of $200 million, are similar to the cost of a one year delay in the repository operation (in 1988 dollars). Thus, if the use of an LSS in the licensing process for the construction authorization can reduce the time period by more than one year, the cost would be justified on that basis alone. Nevertheless, the stated requirements, not cost savings, are clearly the primary justification for LSS.

The predominant cost contribution to LSS is labor, primarily for the capture process and for system operation. The total cost calculations are therefore sensitive to the salaries and the burden rate (the factor applied to salaries to cover fringe benefits, overhead, general and administrative expense, etc.). The calculations presented in this study were based on a burden rate of 2.0 which is typical for a projected dedicated operation of this type. Varying the burden rate to 1.5 would reduce the total life-cycle costs of the Base Conceptual Design to $164 million, while a burden rate of 2.5 would result in a life-cycle cost of $226 million.

The overall conclusion of the costs and benefit study on the eight alternative designs is that the Base Conceptual Design appears to be the preferable design, especially when program and schedule risk is taken into account. However, the results are close enough that this conclusion cannot be dominant, and several other options must continue to be considered.

1) Variant IV, utilizing microfilm images, is only slightly more expensive than the base, due to the higher costs of automated microfilm equipment, while supplying a system which is perceived to be 95% as effective. This would indicate that a search for a more economic automated microfilm system is warranted, along with a more detailed review of the operating characteristics, maintenance, and future viability of this type of system.

2) Variant VI, utilizing re-keying for ASCII conversion has little effect on the user, but is more expensive than a dedicated optical character recognition operation at a re-keying cost of $3.00 per page. The break even point (the point at which the OCR production and re-keying are equal in cost) for re-keying is slightly over $2.00 per page. Thus, the re-keying option should not be dismissed if lower quotes can be obtained.

3) Variant V, utilizing off-line microfilm based production of hardcopy, is also a possible design option provided that copies of documents can be provided to the user overnight.

Certain design implications, however, do appear to be conclusive based on this study:

1) Replicating the system at two locations does not appear to be cost effective since very high system reliability was not a requirement.

2) Hardware full-text search is more costly, offers less benefit, and is a higher perceived risk than software full-text search.

3) Assuming high volumes of hardcopy production, providing images on-line with the associated ability to print limited volumes at the workstation is a cost-effective design.

The sensitivity studies did not provide any evidence that (within the ranges examined) changes in data volume, percent full text, or the number of simultaneous users affected the conclusions derived from the study of the design alternatives. However, it did shed some light on the sensitivity of overall system cost to data volume and percent full text. It would be tempting to conclude that limiting the number of pages to be entered into the LSS or reducing the number of pages available in full text would save millions of dollars. However, it must be kept in mind that these steps would not result in an LSS which either meets the proposed rule or the user needs. The resulting system, while still costing millions of dollars, could run the risk of being not particularly useful for the purpose intended, due to lack of confidence in the user community, and therefore may not contribute to a reduction in the licensing time period.

REFERENCES

DOE, 1987; Licensing Support System Functional Requirements and Design Concepts Report, DOE Office of Civilian Radioactive Waste Management, 31 March 1987.

DOE, 1988a; Licensing Support System Preliminary Needs Analysis, OCRWM, Office of Resource Management, February 1988.

DOE, 1988b; Licensing Support System Preliminary Data Scope Analysis, OWCRM, Office of Resource Management, March 1988.

DOE, 1988c; Licensing Support System Conceptual Design Analysis, OCRWM, Office of Resource Management, May 1988.

NRC, 1987; 52 FR 29024 "Rule on the Submission and Management of Records and Documents Related to the Licensing of a Geologic Repository for the Disposal of High-Level Radioactive Waste; Establishment of an Advisory Committee for Negotiated Rulemaking," 5 August 1987.

APPENDIX A

COST IMPACT OF LICENSING DELAY

COST IMPACT OF LICENSING DELAY

In response to its license application (LA), DOE must obtain a construction authorization (CA) from the Nuclear Regulatory Commission (NRC) before construction can begin at the repository. The Nuclear Waste Policy Act of 1982 levied a requirement that the licensing proceedings for the repository be completed by NRC within three years of their inception. The NRC staff suggested that, using traditional licensing procedures, it would be highly unlikely to meet this schedule and that a period of five to seven years would not be an unreasonable expectation.[1] They regard the LSS as an essential element in reducing the otherwise expected delay in the licensing process.[2]

If the LSS facilitates the suggested two-year reduction in the licensing period, it would not only contribute to compliance with the Act, but benefits of the LSS would be realized as savings yielded by avoiding the costs which would have been incurred as a result of the delay. Significant costs that would be increased as a result of a delay in the period from LA to CA have been identified as follows:

- Development & Evaluation (D&E) costs
- At-reactor storage of spent fuel

Reduction in the licensing delay will thus yield savings by avoiding these costs as discussed below. Table 1 shows the estimated costs of a two-year delay in the duration of the licensing period as well as that of a one-year delay.

Table 1
Estimated Costs of Licensing Delays[a]
(1988 dollars in millions)

	1998	1999	2003 & Beyond	Total
Cost of 2-Year Delay in Licensing Period[b]				
Total Costs				
D&E Costs	$ 60.0	$ 60.0	--	$120.0
At-Reactor Storage	--	--	$275.3	275.3
Total	60.0	60.0	275.3	395.3
Present Value @ 10%	23.1	21.0	43.8	87.9
Present Value @ 3%	44.6	43.4	152.1	240.1
Cost of 1-Year Delay in Licensing Period[b]				
Total Costs				
D&E Costs	60.0	--	--	60.0
At-Reactor Storage	--	--	136.2	136.2
Total	60.0	--	136.2	196.2
Present Value @ 10%	23.1		22.7	45.8
Present Value @ 3%	44.6		76.2	120.8

[a] Benefits of an LSS which would provide a means of avoiding a delay in the licensing period could be measured as the cost savings.

[b] Licensing period refers to the period from license application to construction authorization.

D&E Costs

Development & Evaluation (D&E) is the funding category under which most program costs are currently covered. A work breakdown structure (WBS) has been established by the program to define the activities to be accomplished under these funds. Individual WBS categories were examined in an analysis of impacts of delays on 1986 estimates of program costs by estimating the percentage of effort that would be ongoing in the respective activities during delays in the program schedule including a two-year delay in obtaining the CA, which is scheduled for early 1998.[3,4] Even under the previous program plan, which involved three candidate sites, this was a point when work was scheduled to continue only at the selected site. The stretch-out impacts on the 1986 cost estimates of close to $50 million have been scaled upward to reflect subsequent increases of approximately 50 percent each in 1987 and 1988 estimates of D&E for the selected site during the relevant period.[5,6] The resulting estimate is $110 million for the two-year delay impact which is allocated uniformly at $55 million per year of delay. An additional $5 million per year is estimated for ongoing D&E costs for the monitored retrievable storage (MRS) facility on which work also would be delayed during a delay in the CA for the repository.[7]

The engineering/design costs incurred by the repository project architect-engineer (A-E) and support contractor after LA submittal are separate from the D&E funds. Ongoing costs to maintain minimum staff in these organizations during a licensing delay have not yet been estimated and are not included in the D&E cost impact. These costs are less significant than the estimated D&E costs of delay but represent additional impacts of delay which would increase the benefit of shortening the licensing period.

At-Reactor Storage Costs

A delay in obtaining the CA would delay the completion of construction at the repository by a comparable length of time. This would result in the same delay in the start of operations and the beginning of DOE receipt of spent fuel which is now scheduled for the year 2003. If the receipt of spent fuel from the utilities is delayed, additional storage capacity will have to be established to retain more spent fuel at the reactors which will be continuing to generate additional spent fuel. The timing of the effects at individual reactor sites depends on dates when they are scheduled for spent fuel shipments to DOE. The cost impacts at individual reactor sites are dependent on the remaining capacities in their fuel pools; these costs are for additional facilities to handle the overspill quantities in dry cask storage.

The reference schedule of spent fuel receipts specified in the June 1988 OCRWM Mission Plan Amendment[8] was used as the base from which to measure the additional storage capacity which would be attributable to the slippage in receipts due to the licensing delay. The scheduled receipts of spent fuel were slipped for the duration of the delay. This slipped schedule was compared against the reference schedule to measure the additional quantity of spent fuel to be stored at the reactor sites each year. Using the WASTES computer model,[9] this analysis was done on an individual-reactor basis. This allowed consideration of the spent fuel discharges and inventories and the capacities of the fuel pools at the individual sites to determine how much the additional spent fuel would exceed the capacities of the fuel pools.

The at-reactor-storage cost impact was estimated by assuming that this overspill quantity would go into dry cask storage at the reactor site at an estimated unit cost of about $100/kgU. The cost of adding spent fuel to a fuel pool with available space was treated as being negligible. The 1986 TSLCC analysis addressed the various costs involved in dry cask storage, which consist primarily of the capital costs incurred when added to storage.[10] Escalation of the cask cost, which is the largest element of the costs, resulted in the total unit cost estimate given above. This value falls within the range of unit costs suggested for various storage technologies in a current study.[11]

<u>References</u>

1. William J. Olmstead, September 14-15, 1987. "Alternative Dispute Resolution: The NRC's High-Level Waste Negotiated Rulemaking." <u>Proceedings</u>, American Law Institute-American Bar Association Course of Study, Atomic Energy Licensing and Regulation, page 371.

2. Nuclear Regulatory Commission, August 5, 1987. "Notice of Establishment of an Advisory Committee to Negotiate a Proposed Rule, and Notice of First Meeting." <u>Federal Register</u>, Volume 52, No. 150, page 29024.

3. U.S. DOE OCRWM, April 1986. Analysis of the Total System Life Cycle Cost for the Civilian Radioactive Waste Management Program. Volume 1: The Analysis and Its Results. DOE/RW-0047.

4. Roy F. Weston, Inc., September 1986. Cost Estimating Methods for the Total System Life Cycle Cost Analysis. Prepared for U.S. DOE OCRWM.

5. Roy F. Weston, Inc., September 1987. Cost Estimating Methods for the Total System Life Cycle Cost Analysis. Prepared for U.S. DOE OCRWM.

6. Roy F. Weston, Inc., in preparation, 1988. Cost Estimating Methods for the Total System Life Cycle Cost Analysis. Prepared for U.S. DOE OCRWM.

7. U.S. DOE OCRWM, April 1986, op. cit., page 36.

8. U.S. DOE OCRWM, June 1988. Draft Mission Plan Amendment.

9. Pacific Northwest Laboratories, Battelle Memorial Institute, May 18, 1988 creation date. WASTES computer code, Version 24.

10. Roy F. Weston, Inc., September 1986, op. cit., Chapter 6.

11. Roy F. Weston, Inc. and Pacific Northwest Laboratory, in progress 1988. dry cask storage study.

APPENDIX B

COST ANALYSIS MODEL

The cost calculations and assumptions for the Base Conceptual Design are included in a series of 14 tables which are implemented in Lotus 123. The calculations shown are for the nominal data set (Sensitivity case 111). The tables are:

Table B.1	Workload - Pages
Table B.2	Workload Volume - By Fiscal Year
Table B.3	Workload Volume - This Case
Table B.4	Backlog Processing
Table B.5	Text Conversion Process
Table B.6	Text Conversion Workload
Table B.7	Capture System Process Labor
Table B.8	Capture System Operating Labor
Table B.9	Capture System Process Labor Costs
Table B.10	Capture System Recurring Costs
Table B.11	Capture System Non-Recurring Costs
Table B.12	Search/Image System Operating Labor
Table B.13	Search/Image System Recurring Costs
Table B.14	Search/Image System Non-Recurring Costs

TABLE B.1 WORKLOAD - PAGES

YEAR	LOW ESTIMATE PAGES/YR	LOW ESTIMATE CUM	HIGH ESTIMATE PAGES/YR	HIGH ESTIMATE CUM
PRE 89	--	7,644,000	--	9,685,000
1989	830,000	8,474,000	1,100,000	10,785,000
1990	830,000	9,304,000	1,100,000	11,885,000
1991	1,087,000	10,391,000	1,441,000	13,326,000
1992	1,428,000	11,819,000	1,892,000	15,218,000
1993	1,660,000	13,479,000	2,200,000	17,418,000
1994	2,009,000	15,488,000	2,662,000	20,080,000
1995	1,858,000	17,346,000	2,463,000	22,543,000
1996	1,635,000	18,981,000	2,167,000	24,710,000
1997	1,386,000	20,367,000	1,837,000	26,547,000
1998	1,037,000	21,404,000	1,374,000	27,921,000

Reference: Data Scope Analysis, Table 8
(as corrected 4/26/88)
Additional assumption: 1989 rate same as 1990.

TABLE B.2 WORKLOAD VOLUME - BY FISCAL YEAR

(Table B.1 normalized to fiscal years)

YEAR	LOW ESTIMATE PAGES/FYR	LOW ESTIMATE CUM	HIGH ESTIMATE PAGES/FYR	HIGH ESTIMATE CUM
FY89	830,000	8,266,500	1,100,000	10,510,000
FY90	830,000	9,096,500	1,100,000	11,610,000
FY91	1,022,750	10,119,250	1,355,750	12,965,750
FY92	1,342,750	11,462,000	1,779,250	14,745,000
FY93	1,602,000	13,064,000	2,123,000	16,868,000
FY94	1,921,750	14,985,750	2,546,500	19,414,500
FY95	1,895,750	16,881,500	2,512,750	21,927,250
FY96	1,690,750	18,572,250	2,241,000	24,168,250
FY97	1,448,250	20,020,500	1,919,500	26,087,750
FY98	1,124,250	21,144,750	1,489,750	27,577,500

TABLE B.3 WORKLOAD VOLUME

(Conversion of page estimates from Table B.2 to document estimates)

CASE = HIGH ESTIMATE

YEAR	PAGES PER YEAR	PAGES CUM	DOCUMENTS PER YEAR	DOCUMENTS CUM
PRE 89		9,410,000		1,176,250
FY89	1,100,000	10,510,000	137,500	1,313,750
FY90	1,100,000	11,610,000	137,500	1,451,250
FY91	1,355,750	12,965,750	169,469	1,620,719
FY92	1,779,250	14,745,000	222,406	1,843,125
FY93	2,123,000	16,868,000	265,375	2,108,500
FY94	2,546,500	19,414,500	318,313	2,426,813
FY95	2,512,750	21,927,250	314,094	2,740,906
FY96	2,241,000	24,168,250	280,125	3,021,031
FY97	1,919,500	26,087,750	239,938	3,260,969
FY98	1,489,750	27,577,500	186,219	3,447,188

Assumption: 8.00 pages per document, based on Data Scope Analysis.

TABLE B.4 BACKLOG PROCESSING

(Rate at which workload volume in Table B.3 is processed)

YEAR	PRE FY90 BACKLOG PROCESS	PRE FY90 BACKLOG REMAIN	PAGES/YR	CUM	PAGES/DAY	DOC LOCATION PAGES/YR
FY89	0.0%	100.0%	0	0	0	788,250
FY90	7.5%	92.5%	1,888,250	1,888,250	10,152	3,047,900
FY91	29.0%	63.5%	4,403,650	6,291,900	17,757	2,627,500
FY92	25.0%	38.5%	4,406,750	10,698,650	17,769	2,312,200
FY93	22.0%	16.5%	4,435,200	15,133,850	17,884	1,734,150
FY94	16.5%	0.0%	4,280,650	19,414,500	17,261	0
FY95	0.0%	0.0%	2,512,750	21,927,250	10,132	0
FY96	0.0%	0.0%	2,241,000	24,168,250	9,036	0
FY97	0.0%	0.0%	1,919,500	26,087,750	7,740	0
FY98	0.0%	0.0%	1,489,750	27,577,500	6,007	0

Assumptions:
- 9 months of operation in FY90
- Pages processed = all new pages plus % of backlog
- 248 working days/yr
- Document location pages are the pages from backlog to be processed the following year.

TABLE B.5 TEXT CONVERSION PROCESS

(Distribution of load from Table B.4 among text conversion methods)

YEAR	FRACTION CONVERTED	------------DISTRIBUTION------------ OCR	WORD/PROC	RE-KEYING	TOTAL
PRE 89	95.0%	95.0%	0.0%	5.0%	100.0%
FY89	95.0%	95.0%	5.0%	0.0%	100.0%
FY90	95.0%	75.0%	25.0%	0.0%	100.0%
FY91	95.0%	50.0%	50.0%	0.0%	100.0%
FY92	95.0%	50.0%	50.0%	0.0%	100.0%
FY93	95.0%	25.0%	75.0%	0.0%	100.0%
FY94	95.0%	25.0%	75.0%	0.0%	100.0%
FY95	95.0%	25.0%	75.0%	0.0%	100.0%
FY96	95.0%	25.0%	75.0%	0.0%	100.0%
FY97	95.0%	25.0%	75.0%	0.0%	100.0%
FY98	95.0%	25.0%	75.0%	0.0%	100.0%

Assumptions: 5% of material is not appropriate for text conversion.
Fraction of material submitted in electronic form peaks at 75% in FY93.

TABLE B.6 TEXT CONVERSION WORKLOAD (PAGES/DAY)

(Workload rates in Table B.4 adjusted for distribution in Table B.5 to yield text processing workload)

YEAR	CHAR. RECOG. CURRENT	CHAR. RECOG. BACKLOG	WORD PROCESSING CURRENT	WORD PROCESSING BACKLOG	RE-KEYING CURRENT	RE-KEYING BACKLOG	TOTAL
FY89	0	0	0	0	0	0	0
FY90	4,214	3,825	1,405	21	0	180	9,644
FY91	2,597	11,092	2,597	61	0	523	16,869
FY92	3,408	9,562	3,408	53	0	451	16,881
FY93	2,033	8,414	6,099	46	0	397	16,990
FY94	2,439	6,311	7,316	35	0	297	16,398
FY95	2,406	0	7,219	0	0	0	9,625
FY96	2,146	0	6,438	0	0	0	8,584
FY97	1,838	0	5,515	0	0	0	7,353
FY98	1,427	0	4,280	0	0	0	5,707

TABLE 8.7 CAPTURE SYSTEM PROCESS LABOR (WORKLOAD DEP)

(Staffing requirements for capture process needed to support workload)

YEAR	(1) HEADER ENTRY	(2) DOC PREP	(3) SCAN	(4) MICRO FILM	(5) REBIND	(6) OCR	(7) OCR CORR
FY89	0	0	0	0	0	0	0
FY90	13	13	8	10	4	10	78
FY91	22	22	13	15	6	17	132
FY92	22	22	13	15	6	16	125
FY93	22	22	13	15	6	13	101
FY94	21	21	13	15	6	11	84
FY95	13	13	8	10	4	3	24
FY96	11	11	7	9	3	3	21
FY97	10	10	6	8	3	3	18
FY98	8	8	5	7	2	2	14

CAPTURE SYSTEM PROCESS LABOR (CONT)

YEAR	(8) ASCII INPUT	(9) PAGE CHECK	(10) CATA-LOGING	(11) TROUBLE-SHOOTING	TOTALS CLERICAL (1-6,8,9)	TOTALS EDITING (7)	TOTALS SENIOR (10-11)	(12) SUPER-VISOR
FY89	0	0	0	0	0	0	0	0
FY90	2	2	46	7	62	78	53	13
FY91	3	3	80	11	101	132	91	22
FY92	4	3	80	11	101	125	91	21
FY93	6	3	81	11	100	101	92	20
FY94	7	3	78	11	97	84	89	18
FY95	7	2	46	7	60	24	53	9
FY96	6	2	41	6	52	21	47	8
FY97	5	2	35	5	47	18	40	7
FY98	4	1	28	4	37	14	32	6

Assumptions: (Keyed to columns above)

(1) 15 doc/hr including preliminary duplicate check
(2) 15 doc/hr
(3) 200 pages/hr, 1 operator per scanner
(4) 200 pages/hr, 1 operator per device plus 1 QC per shift
(5) 60 doc/hr
(6) 120 pages/hr for zoning
(7) 15 pages/hr
(8) 20 doc/hr, linking to header and check
(9) 120 doc/hr
(10) 4 doc/hr
(11) 3 doc/hr, 10% document failure rate
(12) 15 production persons per supervisor
1.08 working days/person day

TABLE B.8 CAPTURE SYSTEM OPERATING LABOR (ALL YEARS)

(Capture system labor requirements and costs independent of workload)

Assumptions (based on three capture operating sites):

PERSONNEL	ANNUAL COST EACH	ANNUAL COST EXTENDED
6 system admin persons	$92,000	552,000
9 maintenance technicians	$60,000	540,000
1 software maintenance	$90,000	90,000
1 catalog config mgmt	$80,000	80,000
3 data output technician	$52,000	156,000
6 secretarial/clerical	$36,000	216,000
		$ 1,634,000

TABLE B.9 CAPTURE SYSTEM PROCESS LABOR COSTS $1000

(Costs of capture system labor from Tables B.7 and B.8)

YEAR	CLERICAL	EDITING	SR PERS	SUPER	SYS OPER	LABOR TOTAL
FY89	0	0	0	0	0	0
FY90	1,302	2,808	2,385	1,196	1,634	9,325
FY91	2,828	4,752	5,460	2,024	1,634	16,698
FY92	2,828	4,500	5,460	1,932	1,634	16,354
FY93	2,800	3,636	5,520	1,840	1,634	15,430
FY94	2,716	3,024	5,340	1,656	1,634	14,370
FY95	1,680	864	3,180	828	1,634	8,186
FY96	1,456	756	2,820	736	1,634	7,402
FY97	1,316	648	2,400	644	1,634	6,642
FY98	1,036	504	1,920	552	1,634	5,646

Assumptions:

Annual costs - $28,000 - clerical production staff
$36,000 - editing staff
$60,000 - senior production staff
$92,000 - production supervision

TABLE B.10 CAPTURE SYSTEM RECURRING COSTS $1000

(Labor and non-labor recurring costs for workload rates in Tables B.4 and B.6)

YEAR	# OF STATIONS	DOCUMENT LOCATION	SUBCONT REKEYING	MAINTEN	FACILITY	LABOR	TOTAL
FY89	0	394	0	0	0	0	394
FY90	4	1,524	101	312	380	9,325	11,641
FY91	6	1,314	389	467	570	16,698	19,438
FY92	6	1,156	335	467	570	16,354	18,883
FY93	6	867	295	467	570	15,430	17,629
FY94	6	0	221	467	570	14,370	15,629
FY95	4	0	0	312	380	8,186	8,878
FY96	4	0	0	312	380	7,402	8,094
FY97	3	0	0	234	285	6,642	7,161
FY98	3	0	0	234	285	5,646	6,165
							$113,911

Assumptions:
3000 - pages per day per capture station
10% - per year of capital costs for maintenance
$0.50 - per page in document location costs
$3.00 - per page in rekeying costs
$15.00 - per sqft for space
3000 - sqft per capture station
$50,000 - per capture station per year for facility oper.

TABLE B.11 CAPTURE SYSTEM NONRECURRING COSTS

ITEM	COSTS	EXTENDED
Hardware costs per capture station	$779,000	4,674,000
Facility preparation per sqft	$35	630,000
Software development costs	1,699,000	1,699,000
Software purchase costs per capture station	$27,750	166,500
		$ 7,169,500

Assumptions: Capture system is divided into capture stations.
Capture stations are purchased 2/3 in FY89, 1/3 in FY90.
Six capture stations required

TABLE B.12 SEARCH/IMAGE SYSTEM OPERATING LABOR (ALL YEARS)

(Search/image system labor requirements and costs)

Based on one operating site:

PERSONNEL	ANNUAL COST EACH	ANNUAL COST EXTENDED
1 center manager	$120,000	120,000
4 supervisor systems/operations	$120,000	480,000
5 systems programmers	$100,000	500,000
8 data base admin/prod plan	$110,000	880,000
11 operators	$40,000	440,000
1 admin sec	$40,000	40,000
4 security	$50,000	200,000
4 user support	$70,000	280,000
		2,940,000

TABLE B.13 SEARCH/IMAGE SYSTEM RECURRING COSTS (in thousands)

(Search/image system labor and non-labor operating costs)

YEAR	LABOR	MAINTEN	FACILITY	TELECOM	HARDCOPY	TOTAL
FY89	0	0	0	0	0	0
FY90	0	0	0	0	0	0
FY91	2,940	823	890	360	542	5,555
FY92	2,940	1,175	890	360	712	6,076
FY93	2,940	1,466	890	360	849	6,506
FY94	2,940	1,674	890	360	1,019	6,882
FY95	2,940	1,857	890	360	1,005	7,053
FY96	2,940	1,857	890	360	896	6,944
FY97	2,940	1,857	890	360	768	6,815
FY98	2,940	1,857	890	360	596	6,643
						$52,474

Assumptions: Operation begins in FY91 (Oct 90)
10% - per year of capital costs for maintenance
$890,000 - per year for facilities operation
$360,000 - per year for operation of communications
4 - ratio of hardcopy pages to pages produced
$0.10 - per page to produce and ship hardcopy

TABLE B.14 SEARCH/IMAGE SYSTEM NONRECURRING COSTS

HARDWARE:	Search system	$ 6,800,000	FY90,91
	Search system storage	7,355,000	FY90,92,93
	Image system	343,000	FY90
	Image system storage	1,691,000	FY90,92
	Hardcopy production	238,000	FY90,91
	Telecommunications	475,000	FY90
	Workstations	678,000	FY90,92,93

	Subtotal	$17,580,000	
SOFTWARE:	Purchased - Search system	708,000	FY90
	Purchased - Image system	0	
	Purchased - Workstations	21,580	FY90,92,93
	Developed - FY89	816,000	FY89
	Developed - FY90	1,226,000	FY90

	Subtotal	$ 2,771,580	
FACILITY:	Preparation	$ 1,060,000	FY89,90
	Total	$21,411,580	

Assumptions: All expenditures in FY as noted
Hardware costs include installation

APPENDIX C

DERIVATION OF HARDCOPY ESTIMATES

Since there was no source of directly comparable experience from which to obtain estimates of the number of hardcopy pages that would be requested, three independent estimation methods were used and the results compared. Each of the methods were based on the expected maximum use period and the user behavior associated with text retrieval systems with images available on-line, but hardcopy still very much in use. This desire for hardcopy was identified in the Preliminary Needs Analysis.

The primary purpose of the estimates was to determine the printer capacity required to support the production and distribution of the requested copies. Thus the estimates were developed to minimize the probability of under sizing the hardware.

The first method was based on a concept of parties to the licensing process building hardcopy libraries to support their cases. The factors in this method are:

The number of parties = 3 majors and all others = 0.3 total = 3.3

The peak data base size = 27,000,000 pages in 1998 (DOE, 1988b)

The percent of pages in the LSS that are not already conveniently available to a party is 80% of the data base.

The percent of the above pages that are of interest = 10%

The fraction of the pages of interest that would be requested for remote printing and shipping in any one year = 1/6

The average number of working copies a party would need (multiple staff members retaining copies) = 10

In equation form:

$$3.3 \times 27{,}000{,}000 \times 0.8 \times 0.1 \times 1/6 \times 10 = 10{,}600{,}000 \text{ pages/year}$$

The second method was based on the number of users and the number of pages requested each day by each user. These estimates where then used in the following formula to obtain the number of pages in a peak year.

Pages/year = # of users x pages/user/day x days/year

The best estimate of the maximum number of users from the Preliminary Needs Analysis is 375.

During a peak use period, the number of pages requested per day was estimated to be 150 per user.

Number of working days/year is 240.

375 x 150 x 240 = 13,500,000 pages/year

The third method is based on the National Technical Information Service (NTIS) experience, on a data base of comparable size. A comparison was made between NTIS experience and projected LSS usage, considering the annual number of sessions, the fact that images are not available on-line at NTIS, and the difference in average document size. On the basis of this comparison, it was estimated that the LSS would print only 20% of the amount of pages that is requested from NTIS each year.

0.2 x NTIS = 11,000,000 pages/year

All three independent estimates are very close, each being about 10,000,000 pages/year, for a peak year. To estimate the number of hardcopy pages in any year, the peak year level was determined from the activity histogram in the Preliminary Needs Analysis. 1994 and 1995 are shown to be the peak use years. The average number of pages added in 1994 and 1995 is about 2,500,000, which is also proportional to the activity histogram. This yields a factor of 4 for the number of hardcopy pages requested as a function of the number added. This factor was used in the cost model in the base case.

In the cost model, the number of hardcopy pages per year was used to compute the operational cost of producing and distribution the hardcopy. Based on the analysis in Section 4.2, the hardcopy distribution costs are a very small percentage of the total, indicating that the number of hardcopy pages is not an important factor in the total cost.