109
Received with letter dtd 9/27/88

Office of Civilian Radioactive Waste Management

Office of Resource Management

Licensing Support System Benefit-Cost Analysis

July 1988

U.S. Department of Energy
Office of Civilian Radioactive Waste Management

Office of Civilian Radioactive Waste Management

Office of Resource Management

Licensing Support System Benefit-Cost Analysis

July 1988

U.S. Department of Energy
Office of Civilian Radioactive Waste Management

LICENSING SUPPORT SYSTEM
BENEFIT-COST ANALYSIS

TABLE OF CONTENTS

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

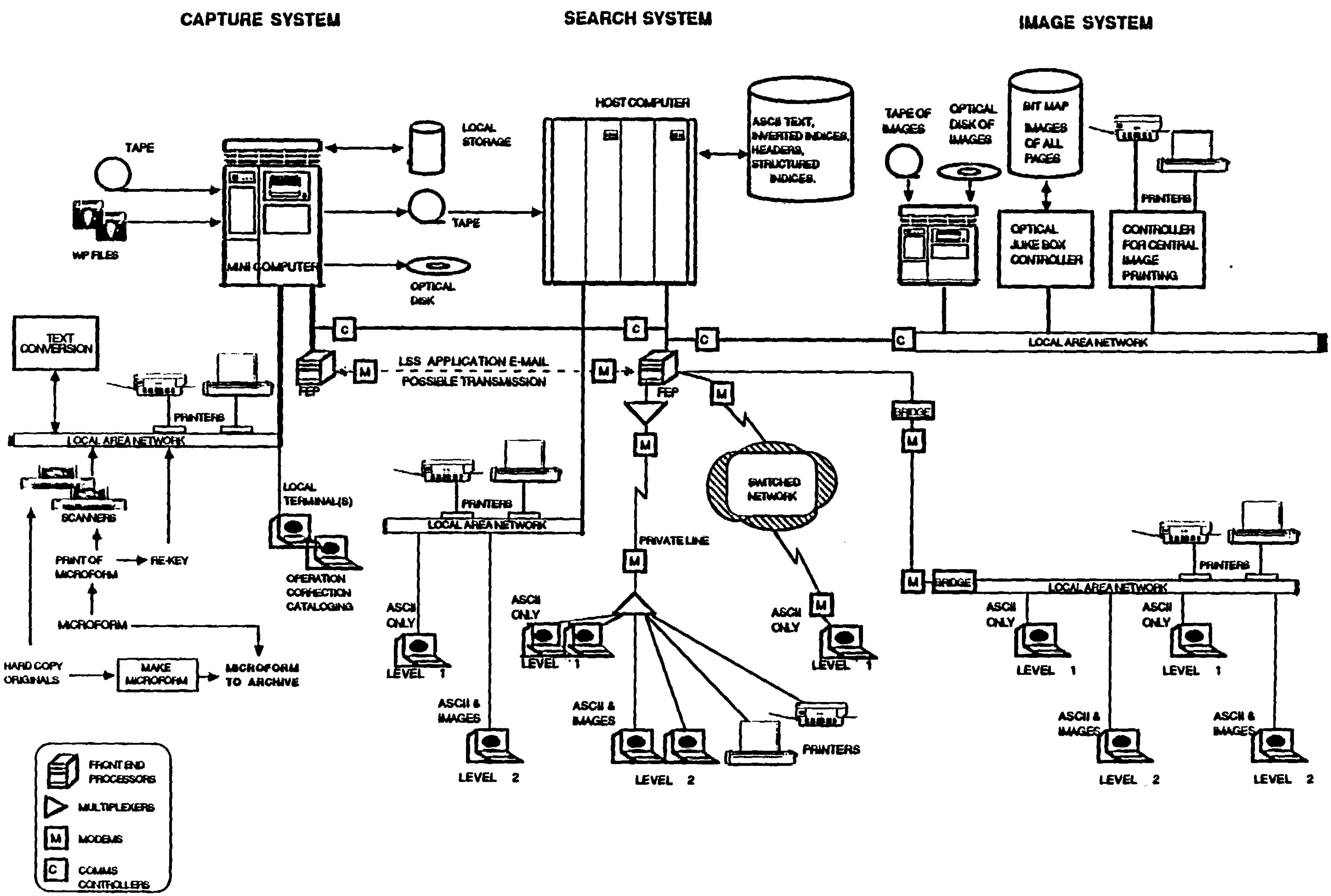
The capture system consists of the operating procedures, computer, image and character recognition equipment, software and facility to process documents submitted for entry into the LSS. The capture station processes are:

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

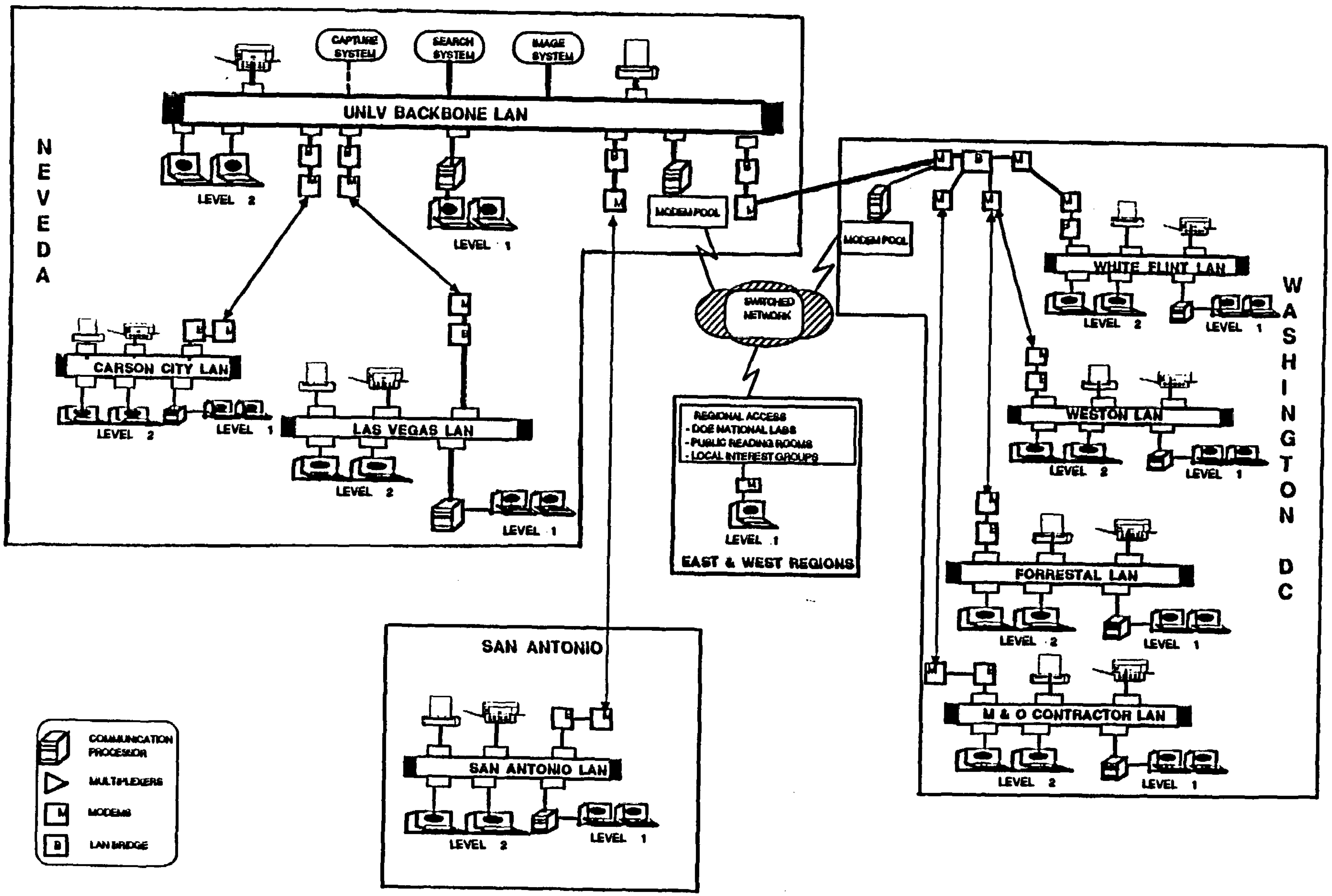
User queries and display requests, electronic mail messages, and ASCII text and image data to be displayed or printed at the workstations are transmitted over the LSS communications system.

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.

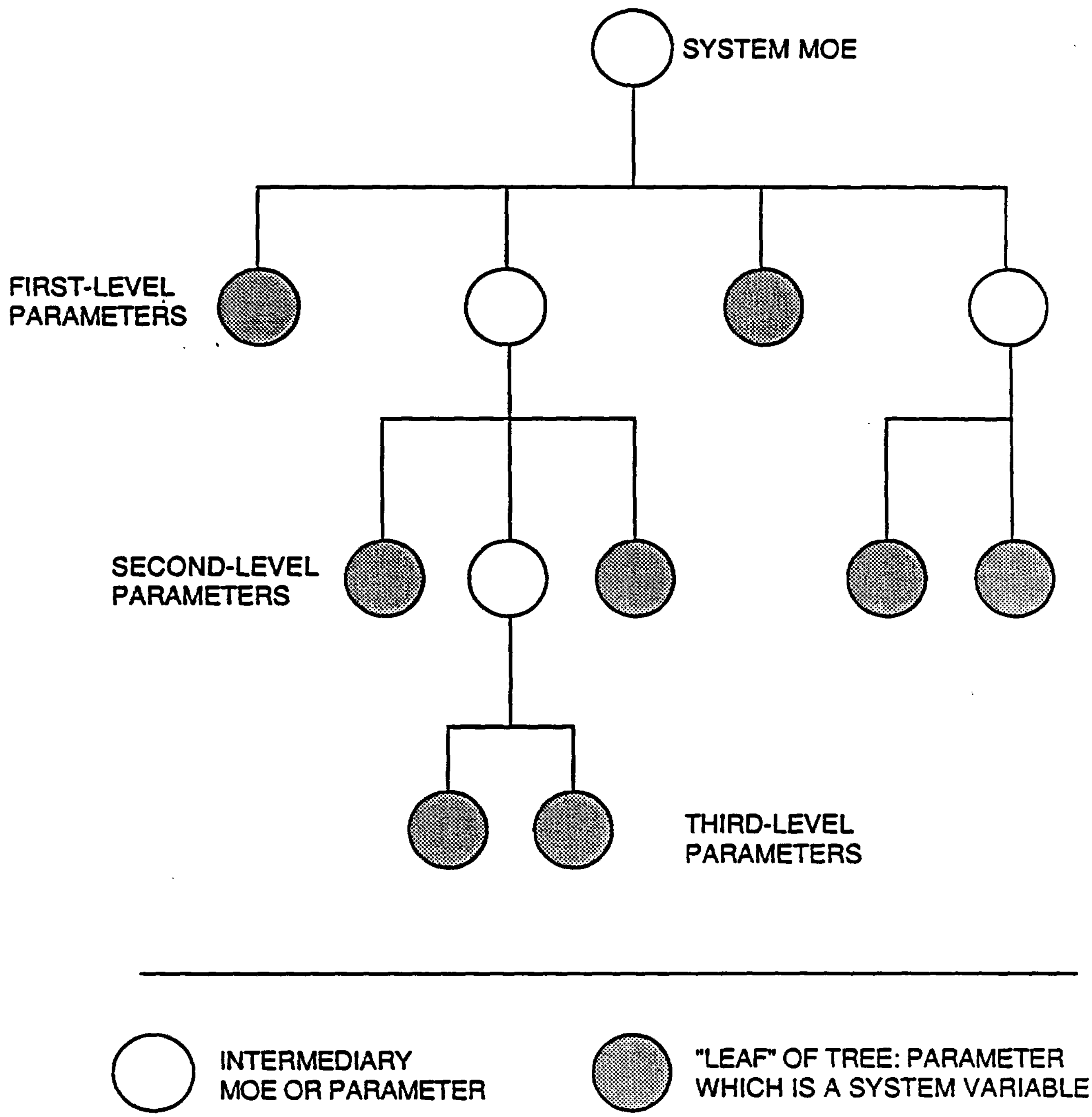


Figure 3. Tree Evaluation Parameters

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.

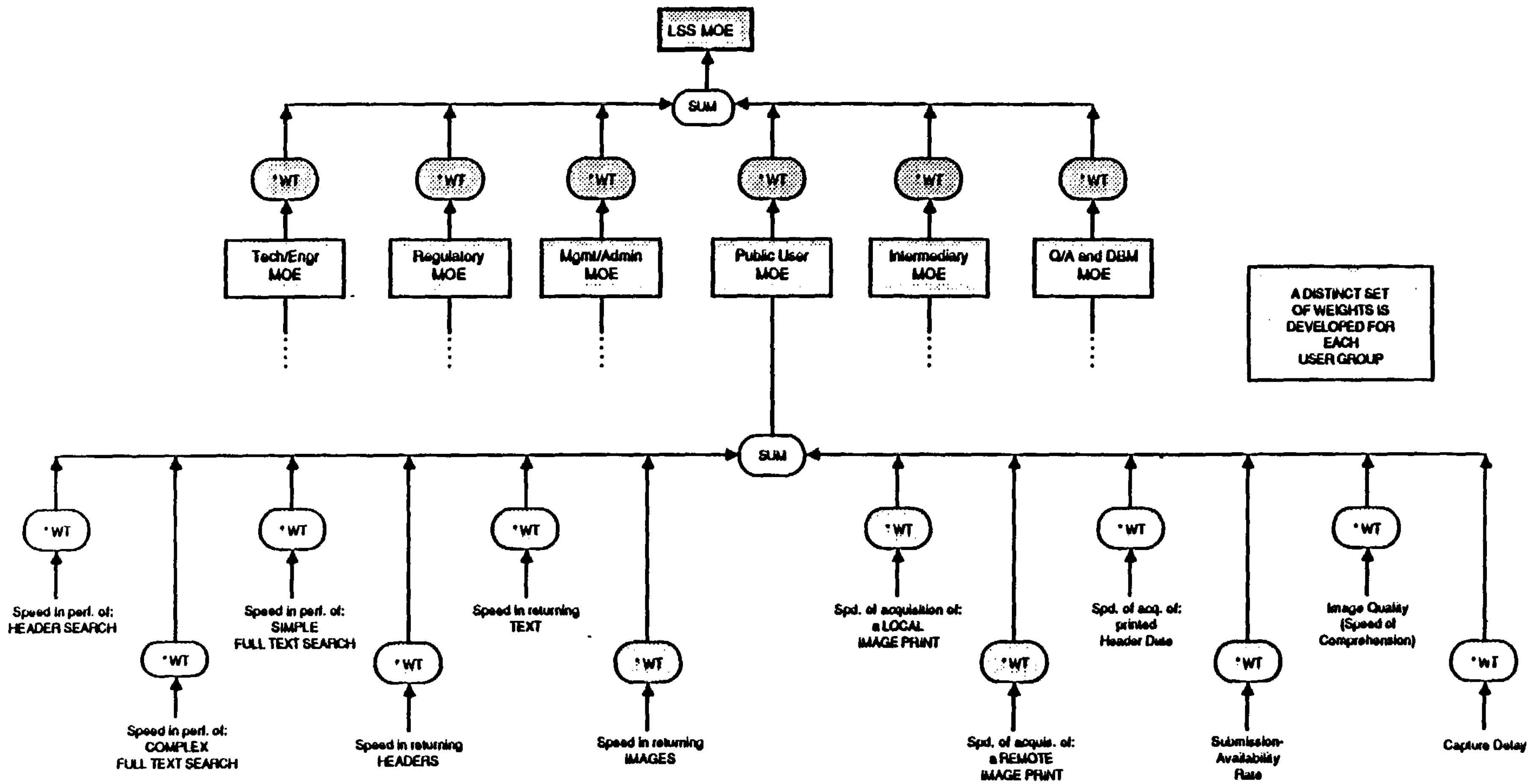


Figure 4. LSS Benefit Analysis Tree

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

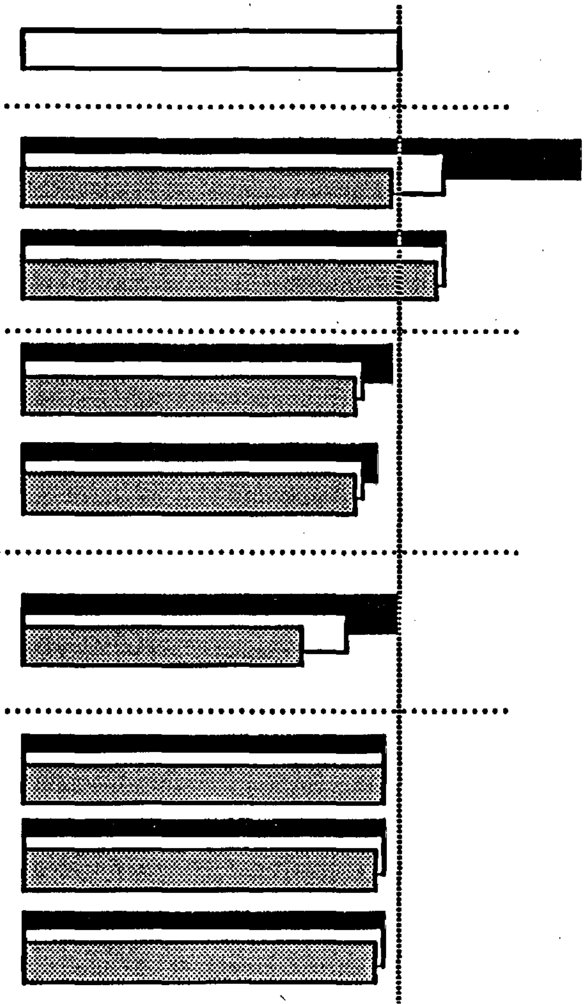


Figure 6. LSS BENEFITS SUMMARY AND SENSITIVITY ANALYSIS

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

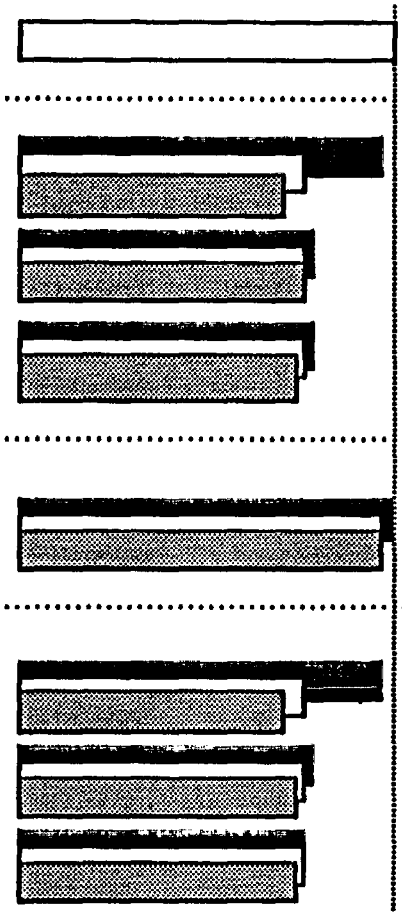


Figure 6. LSS BENEFITS SUMMARY AND SENSITIVITY ANALYSIS (Continued)

5.2.2.3 Variant II - Hardware Full-Text Search

In the Base Conceptual Design full-text search is implemented 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. All other functions (E-mail, header searches, etc.) 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]

References

1. William J. Olmstead, September 14-15, 1987. "Alternative Dispute Resolution: The NRC's High-Level Waste Negotiated Rulemaking." Proceedings, 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." Federal Register, 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

	PAGES		DOCUMENTS	
YEAR	PER YEAR	CUM	PER YEAR	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)

	PRE FY90 BACKLOG					DOC LOCATION
YEAR	PROCESS	REMAIN	PAGES/YR	CUM	PAGES/DAY	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 B.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)	EDITING (7)	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	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 x 27,000,000 x 0.8 x 0.1 x 1/6 x 10 = 10,600,000 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.