

# History of LSN Functional Requirements

## 1 Current Status

### 1.1 Rule Directed Functionality to Be Provided by The Participant Systems.

Each participant must meet core requirements related to making their documentary materials available and for providing the computerization necessary to comply with the requirements identified in the rule for document production and service. These include requirements for providing structured data bibliographic headers and searchable text of its documentary material, and, providing a statement of where an authenticated image of the document may be obtained by users. Where text is not available, the image must be made available online in lieu of the text file. Structured data bibliographic headers, only, are required for items not suitable for image or text. Similarly, structured data bibliographic headers, only, are required for privileged, confidential, safeguards and other types of limited access documents.

### 1.2 Functionality Associated With NRC's Docket in ADAMS.

NRC's SECY organization, in its management of the official docket, must provide a docket which receives, stores, distributes, and maintains pre-license application docket materials. In addition to the official docket, there is a requirement to provide a Protective Order File. Unavailability of the electronic docket for more than 4 hours in any day must be communicated back to the presiding officer so that the day is not counted in the computation of time. Case management capabilities associated with the docket include transcript and deposition exhibit management.

### 1.3 Functionality Associated With NRC's Electronic Information Exchange Systems.

Each participant must utilize a secured, electronic process by which all filings are able to be submitted/received electronically to comply with service requirements.

### 1.4 Procedural or External System Capabilities.

Hardware and software capabilities are augmented with procedural requirements outlined in the rule. Participants must demonstrate substantial and timely compliance with these procedural requirements in order to be granted party status. Key agreements include the requirement that participants follow document and data format standards for providing electronic access and the requirement that they follow procedures and standards for motions practice.

Participants must designate an official who is responsible for the administration of making their documents available. This includes establishing procedures to make their documents available, to ensure that each of their documents has a unique ID, to train their own staff on how to make their documents available, etc. The rule also requires participants to have a designated official certify that procedures have been implemented, that documentary material has been made electronically available, and, update these certifications at 12 month intervals.

DOE, as participant must additionally submit the license application to the docket in electronic form.<sup>1</sup>

## **2.0 Requirements Development**

This section tracks the history of functional requirements associated with the system intended to meet the discovery requirements outlined in 10 CFR Part 2, Subpart J. It then presents the high level and secondary level functional requirements associated with fulfilling this mission.

### **2.1 Background**

Prior to the promulgation of 10 CFR Part 2, Subpart J, a number of studies were conducted by both the U.S. Department of Energy and the U.S. Nuclear Regulatory Commission. These studies were used to define the mission for the originally conceived Licensing Support System and were reflected in the negotiations conducted to formulate the rule.

### **2.2 Functionalities Considered During Rulemaking**

Preliminary needs assessments by both NRC and DOE and their contractors generated fairly consistent lists of system requirements prior to the completion of the negotiated rulemaking process. The following list is not exhaustive but does indicate the thinking of both of these agencies going into the rulemaking process.

- The system must be interactive and comprehensive with regard to licensing information (i.e., it must be a useful tool for all participants in the licensing process.<sup>2</sup>
- The system must provide rapid access to the information regardless of the user's geographic location and the geographic location of the computer system containing the information.
- The record must be durable and extend for 60 to 90 years, until repository closure.
- It must provide a comprehensive reference source of all regulations and other regulatory guidance documents applicable to the repository licensing process.
- It must maintain the confidentiality of any information which must have such protection.
- It must provide rapid search and full text storage, searching and retrieval.

---

<sup>1</sup> 10 CFR Part 2, Subpart J, §1013(a)(2)

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

These studies included analysis by a nationally recognized litigation support information management consultant.<sup>3</sup> To the initial DOE analysis, the following needs were identified and verbalized:

- Security: the system must have the ability to assure restricted access to specifically identified information.
- Retrievability: the retrievability of information in the system must be extremely high, at or near the state of the art, for it to be acceptable for legal discovery.
- Recall: the system must be able to deliver a high degree of recall, e.g., the proportion of documents found in a search, relative to the total number of documents in the database which meet a search criteria.
- Precision: the system must be able to deliver a high degree of precision, e.g, the proportion of truly relevant documents found in a search relative to the total number of documents found.
- Speed: the size of the data base dictates that for the system to be of practical use during discovery, the system must make use of state-of-the-art technology to maximize access and minimize delay time in identifying material sought.

This perspective was added to another study conducted by another NRC contractor, Aerospace, Inc.<sup>4</sup> They identified the following list of needs, some of which overlaid on the work performed by Weston and Jordan. Again, this is not the exhaustive list, but the key themes are included:

- Comprehensive Content: Store in retrievable manner the full text of records that meet the relevancy and topical guidelines of the Rule.
- Broad Indexing Capability: be able to search and select by keywords, or descriptor phrases, that define the subject, author, and title by significant words in context in abstracts and text. Be able also to search on date, issuing agency, identifying number, and other necessary identifiers. To ensure precision and recall, the search routine must operate on both standard abstract terms (e.g., title, author, keywords) and the full text.
- Prompt Response: verify the existence of a record, determine the location of a record, and display, on line, the full text of records resident in the database in real time at authorized user terminals. Produce hard copy of any record by use of local laser printing.

---

<sup>3</sup> John S. Jordan & Associates. Discovery and Rulemaking Perspectives on the Use of an Information Storage and Retrieval System in the Licensing Proceedings for the High-Level Waste Repository, prepared for the U.S. Nuclear Regulatory Commission. February 14, 1986.

<sup>4</sup> Aerospace Corporation. Requirements Definition for Licensing Information Management System for Nuclear Waste, Draft Revision 2, Prepared for the U.S. Nuclear Regulatory Commission, March 1986.

- Operational Availability: make the system available as soon as possible for the precensing phase and have it remain operational during licensing and after licensing.
- Security: Protect against the loss and destruction of records and protect privileged material by controlled access.
- Simplicity of Use: Provide access for non-technical users with no prior online data base interaction experience.
- Long Term Viability: Incorporate the capability to improve service throughout the licensing period.
- Accessibility: Provide remote access to the database is accessible to personnel of the States, Indian tribal organizations, and the general public.
- Compatibility: Provide interfacing capability to other information management system and data bases.
- Reliability: Maintain a system with minimum downtime. Protect data files during system crashes and provide recovery in less than 24 hours.
- Affordability: Provide a system that is cost effective.
- Standardization: Use data transmission, library and information science standards, and other industry data and hardware and software standards for records transfer, date exchange, etc.
- Completeness: All participating parties must acknowledge that they have adopted and followed procedures ensuring that all relevant records are submitted to the system.

### **2.3 Functionalities Reflected in the LSS Rule**

These needs translated into the following functional requirements which were aggregated for DOE by yet another contractor.<sup>5</sup> These intended functionalities were represented in the final version of the negotiated rule on the proceedings. Again, this list is somewhat abbreviated from the original:

- Structured index searching.
- Full text searching.
- Search and Retrieval interface should be GUI.

---

<sup>5</sup> Arthur Young International. LSS Functional Requirements and Design Concept Report; Prepared for OCRWM, March 31, 1987.

- The search criteria during a query should be able to be retained.
- User help should be available online.
- Bit mapped images of documents should be available as well as text.
- Summary descriptions of retrieval information and counts of retrieved items should be available.
- The system procedures should verify the integrity of the information at the time it is captured into the system.
- The system procedures should verify that the information has been authorized for entry at the time of capture.
- The system should avoid duplicate records.
- The information in the system should be verified for conformity with the original source document.
- The association between cataloged information and the actual content of the document should be verified.
- Access to the system should be controlled and restricted to authorized users, including members of the public.
- Accidental or malicious destruction and alteration of information in the system should be prevented.
- The system should be able to detect and prevent “unreasonable” queries which would overburden the system.
- The system should support display of documents either in text or bit-mapped image format, and should be able to highlight search information and other items of special interest.
- The system should be able to display both cataloged and full text information.
- The system should be able to deliver text for downloading.
- The system should be able to support multi-user access from around the country.
- The system should provide suitable backup and recovery capabilities.

## **2.4 Functionalities Entering LSS Design Phase**

After rulemaking, the system was described as follows in a formal DOE<sup>6</sup> document:

- The system should include capabilities for managing various types of databases.
- Information should be stored in the form of headers for all records and in full text for many, if not all, documents.
- The system should include the capability for efficient and accurate data retrieval using a variety of methods.
- Data should be capable of being reduced to hardcopy form by the users.
- The system should be capable of generating various types of reports.
- The system should be easy to use with minimum training necessary, containing built-in help functions, and providing assistance when needed, either through an expert system or on-call assistance.
- There should be a procedure to identify and minimize or avoid duplicate records.
- The records should be maintained in a secure environment.
- Electronic information exchange capabilities should be provided.
- The data should be entered and maintained under an independently verified quality assurance program.
- Structured index searching via detailed and extensive headers should be available, involving subject terms and keywords assigned with the aid of a controlled vocabulary.
- Full text search capability on both document text and headers should be available.
- The system should include data bases for document, regulations, tracking of issues and commitments, and indexes for non document materials such as physical sample inventories.
- Search aids including thesaurus, boolean logic, and proximity searches should be available to meet the performance requirements identified.
- The number of users needing the system will exceed 350 at peak demand.
- Major geographic centers of users will be Washington DC, Las Vegas, Nevada, and San Antonio Texas. Other locations, comprising about 10% of the users, are expected.

---

<sup>6</sup> U.S. Department of Energy Office of Civilian Radioactive Waste Management. Licensing Support System Preliminary Needs Analysis (DOE/RW-0209), January 1989.

- Discernible usage patterns will exist in terms of user characteristics, time-of-day use, and use in relation to the program schedule.
- The system should be designed to maximize recall and precision with minimum performance targets on the order of 80 percent.
- Response to large queries should be available in approximately 15 to 30 minutes.
- Documents should be available for viewing on a screen in less than 30 seconds after they are identified.
- Electronic (bit-map) images may be required for some records, particularly graphs, diagrams and maps.
- The capability for downloading data may be desirable.
- Some files such as privileged information or annotated records may require limited access or privacy.
- Some users have expressed interest in a priority access system.

## **2.5 Functionalities Refined During LSS Design**

The LSS design functional requirements were significantly refined by the LSSARP Technical Working Group and forwarded to the ARP members for consideration on May 17, 1995. The LSSARP did not formally vote on the design phase functional requirements at that time. Subsequent to these functional requirements being developed by the DOE support contractor, the OCRWM program delayed the license application schedule, and the original LSS rule in 10 CFR Part 2, Subpart J was revised to allow the utilization of Internet-based software to fulfill the discovery mission earlier defined and agreed to during the negotiation of the original LSS rule. The following section details the 1995 functional requirements with commentary on their applicability with regard to use of Internet-based technology solutions.

## **3.0 Current Functional Requirements**

Licensing Support Network (LSN) functional requirements are derived from a series of statements that are found in 10 CFR 2, Subpart J. Capabilities addressed in the rule include:

- 1) Internet-based text search and retrieval capability;
- 2) functionality to be provided by the participants' systems independent of capabilities of the LSN;
- 3) adjunct capabilities of the NRC's separate electronic docket;

- 4) adjunct capabilities of NRC's separate electronic information exchange system;  
and,
- 5) procedural direction external to the LSN but which facilitate its uniform operation.

The specific method of providing access to documentary material is not mandated by the LSN Rule in order to accommodate on-going and future technology advances. [63 FR 71735, SUPPLEMENTARY INFORMATION]. For the purposes of developing these requirements, the technology to be adopted for development and implementation of the LSN is based on the facility popularly known as the "World Wide Web" (WWW or web) on the Internet. More specifically, this can be defined as dissemination of information with HTTP (HyperText Transport Protocol) servers to HTTP clients.

The requirements described in 10 CFR 2, Subpart J are subject to subsequent guidance or clarification insofar as they are only broad directives on what the system is intended to do. Additionally, these statements are strongly influenced by the technology solutions that are subsequently applied to the problems. A detailed statement of the functional requirements for the system is included in APPENDIX A.

### **3.1 Functionality for Docket System**

The docket system must, first and foremost, incorporate all of the underlying system requirements which reflect NRC's docket and records management directives and the established procedures and standards for motions practice. For the Yucca Mountain Repository licensing, the docket functionality begins 30 days after DOE submission of the LA to the President. The LSN utilizes NRC's existing docket structure to provide a docket which receives, stores, distributes, and maintains pre-license application docket materials. It is a fully featured capability, and allows for creating bibliographic headers for transcripts (and associated exhibit materials) submitted which were not previously made available in a participant collection. It also allows for creating a digital image of each page of electronically submitted text material (depositions, transcripts, attached exhibits) which were not previously made available as an image in a participant collection. The software environment in with the docket resides preserves transcript formats including page and line numbers in the image format for uniformity in reference. It also provides the ability to link document records in the docket file with their reference in full text transcript. Additionally, it already includes the ability to display limited access warnings.

For authorized individuals (SECY), NRC's Docket:

- provides a mechanism by which transcripts may be entered to the docket on a daily basis in order to provide next-day availability at the hearing;
- provides mechanisms for participants to transmit depositions to NRC (SECY) in electronic form for entry into the docket; it also provides a mechanism to authorized users (SECY) to identify only a part of a deposition which is offered into evidence;

- provides an electronic list of all exhibits with a way of indicating where they were introduced; and
- is the repository for the protective order file materials.

NRC's SECY organization is authorized to maintain the docket. The LSN site will contain a link directly into the existing NRC docket files and passworded access to the protective order file. This is one (of a number - such as direct access to ADAMS external server) mechanisms by which the presiding officer and all counsel of all parties have access to the electronic docket (including protective order file) during the hearing.

### **3.1.1 Docket Submission Controls**

The docket must implement procedures and tracking mechanisms to allow document submitters to verify that document information entered into the LSN docket database is identical to the document information submitted to the LSN. Additionally, the docket environment must implement certain data management capabilities, including:

- a mechanism preserve transcript formats including page and line numbers in the image format for uniformity in reference; and,
- the ability to link document records in the docket file with their reference in full text transcript. *The implementation of this should not use hyperlinks (records retirement issues) and should be accomplished in single environment that is easily understandable and quick to learn.*

### **3.1.2 Docketing Process**

Each participant is required to provide the computer system necessary to comply with service requirements (Rule 2.1011(b)). This includes providing an automated/procedural mechanism to show a proof of service for each document filed into the docket (Rule 2.1013(c)(4)) to the extent of capturing the event of a document being placed in a recipient's mailbox in order provide proof of the completion of service (Rule 2.1013(c)(3)). *This may be accomplished by procedurally requiring an echo response to incoming e-mail if received from a defined party to the licensing.* The LSN docket must also implement procedures and tracking mechanisms to track the receipt (date, time, accepted/rejected status) of documents submitted to NRC from receipt through disposition.

### **3.1.3 Docket Availability**

The LSN docket environment must provide a mechanism that allows an authorized user (SECY) to identify the unavailability of the electronic docket for more than 4 hours in any day, and a means to communicate unavailability to the presiding officer so that the day is not counted in the computation of time.

## **3.2 Availability of electronic data**

### **3.2.1 Integrity of Data**

### 3.2.2 Audit Tools

The Audit system must provide software tools to:

- collect audit data on all aspects of LSN performance;
- store audit data as a record;
- aggregate, compile, crosscut, and otherwise analyze audit data by providing a software capability (package) which can be used to perform the analyses and provide mechanisms to generate output and reports;
- create user-defined reports which include descriptive statistical computations and sub-grouping (ad hoc);
- store user-defined, statistical reports on participant, portal, and docket system activities; and,
- print statistical reports on participant, LSN server, and docket system activities.

The Audit system should allow only authorized individuals to enable/disable the audit functions and to backup and remove audit files from the system. Additionally, the audit server should adhere to the same standards, where applicable, as the LSN search and retrieval server in order to ensure compatibility.

## Discussion of LSN Design Alternatives

we are presenting an in-depth analysis of three alternatives, designated CPIC #1, CPIC #2, and CPIC #3 as indicated in the following table.

In the first proposed design, search and retrieval requirements are met by the participants individually; in the second and third designs these requirements are met by the LSN site. A total of five alternative solutions were evaluated by the Licensing Support Network Advisory Panel's Technical Working Group. The Licensing Support Network Advisory Review Panel (LSNARP) met in Las Vegas, Nevada, on February 23, 2000, to consider various alternative computer system architectures for the Licensing Support Network (LSN).

At the February 23, 2000 ARP meeting, the Panel received the report of its Technical Working Group (TWG). That group, which is composed of computer technology representatives of the Panel's members, had been chartered at the first LSNARP meeting in October 1999 to examine in depth potential alternative computer system solutions and report back to the Panel. The TWG studied and presented five alternatives as noted on the charts on following pages.

<b>Technical Working Group &amp; Recommendations (✓)</b>	<b>ASLBP Recommended for CPIC Characterization Based on Participant Input</b>	<b>CPIC Alternative Designation</b>
Alternative 1 - Simplified - Linking Web Pages	(✓) Strong Participant Interest	<b>CPIC # 1</b>
Alternative 2 - Moderate - Centralized Search Interface Serviced by Participant Sites	No Interest Expressed by LSNARP	
(✓) Alternative 3 - Portal - Comprehensive Interface with Files Delivered by Participant Sites	(✓) LSN Administrator Recommendation & Some Participant Interest	<b>CPIC #2</b>
(✓) Alternative 4 - Same as Alternative 3 But All Participant Machines Co-located	Consensus ARP Vote Against	
(✓) Alternative 5 - Same as Alternative 3 But Participant Files Copied to a Central Store Operated by LSN	(✓) Some Participant Interest	<b>CPIC #3</b>

The LSN can be regarded as consisting of three functional components. Specifically, these are:

- A component that aids the LSNA in auditing participant compliance with the LSN Rule.

- A component that presents LSN information to participants, other interested parties, and the general public.
- A component that stores LSN documentary information for the use of components one and two.

### **1.9.1 CPIC Alternative #1 - Participant Delivered Capabilities Linked by the Web**

### **1.9.2 CPIC Alternative #2 - Portal with Comprehensive Interface with Files Delivered by Participant Sites**

This design alternative was validated by the TWG and has been named the “distributed alternative” by the members of the TWG. This terminology focuses on a key characteristic of the design - that each participant, assemble, prepare, and publish their own collections of documents on a WWW server that they control at a site of their choosing. It is an analogue of many other WWW systems extant on the Internet, such as Yahoo or Lycos, and contains many of the same components as these analogous (but much larger) systems.

The following sections describe the details of its components, how they “fit” together, the hardware and software used in the design, and an approximation of the component life-cycle costs.

#### **Compliance Component**

This component is used by a small set of LSNA users who require specific information at specific times to fulfill data assuredness requirements. It provides a mechanism to scrutinize the effectiveness of the LSN, to ensure that it is functioning as intended to the end-user base. Many functions of this subsystem will be accessed through a WWW browser using standard HTTP/HTML mechanisms bolstered by CGI programs. It includes a data retrieval element that routinely “roves” participant sites, fetching participant data (documents, statistics, and other) and storing this data pending processing. The primary method of following the operation and evolution of the LSN is through reporting mechanisms and this component provides the necessary reports on LSN functionality to enable the LSNA to ascertain participant compliance and aid in determining whether remedial action is required. Reports will be generated automatically by the system on a periodic basis, when exceptional conditions arise, and on-demand. The full array of required reports is yet to be determined. Minimally required are: a) listings of changes in participant document collections, i.e. additions, deletions, and modifications, and, b) the “health” of the LSN, component and sub-component uptime and performance data (e.g. web server hits, average response times, number of users, etc.). Historical trends will be maintained both in HTML tables and graphically. Collected data is stored and processed, and the results of that processing are likewise stored to build a file for trend analysis. The database will be a network-capable SQL relational database that will provide structured data to both front ends, i.e. the compliance and presentation components. Both file system storage and database storage will be accommodated. It also gathers low-level statistics on network operation for trend and throughput analysis.

This element provides a level of assuredness that the systems housing the LSN are functioning as required, specifically validating security mechanisms, network monitoring to evaluate hardware and software outages or sub-optimal operation, and physical plant including environmental and power conditioning. This sub-element provides appropriate and implements disaster recovery mechanisms, e.g. a backup/restore capability.

A single computer system of the workstation class is adequate for this functionality. The security sub-element mandates that the system be separate from and more restricted than the computer system (described below) that provides general access. The system should be equipped with the standard components, a graphical display, and a device appropriate for backup. The following software components will be required: a web server (e.g. Apache, Netscape Enterprise, MS IIS), a database with accompanying report generation software (e.g. PostgreSQL, Oracle, MS SQL Server), firewall software (e.g. IPFW, ipfilter, Firewall-1), network monitoring and management software (e.g. Big Brother, SunNet Manager, HP OpenView), and a web spider (e.g. MoMspider, BRS/Search, Fulcrum Search Server). Note that the web server, database, and web spider are also part of the presentation component described below). The same software can be used for both purposes. In addition, it is anticipated that this component will require some custom software, scripts and CGI's rather than full-blown applications.

### **Presentation Component**

This component is a "front-end" with a large set of users who require access to a wide range of information at arbitrary times. It is intended to fulfill the requirement to provide search and retrieval and information delivery to the parties through WWW technology. The intended user base includes all participants and potential participants, the LSNA and his designees, the press, and the general public.

It is a WWW presentation interface with additional sub-components that consist of:

- Introductory and overview documentation.
- Training / tutorial materials on how to use the site to obtain LSN-related information, and the other aspects of the site, and how to submit to the docket.
- Portal software that allows user customization of user interfaces and user document search and access strategies.
- A search facility that allows LSN-wide searching of participant materials, including individual user custom searching strategies.
- Publication of statistical information on LSN participant sites, including site content and performance.
- A web-based interactive forum in which interested parties can discuss or exchange information regarding LSN matters.
- Help-desk assistance (with escalation) for participants and public users.

- A LISTSERV (e-mail list manager) to allow participants to easily send electronic mail to all interested parties. A number of mailing lists will be created as needed for discussion of specific subjects, including a list with the e-mail addresses of all participants for notification purposes. The LISTSERV software will allow each participant to manage their own subscriptions to interest lists and archive messages to the lists. The LSN is not intended to provide a public LISTSERV function.
- Aggregation and publication of overall LSN access and usage statistics, e.g., number of hits.

Web browsers will be the predominant access method to this component. It is anticipated that this will be the sole access method for the majority of users. Browsers will be used to gain access to general information, participant documentary collections, and to discussion forums.

A single server class computer system comprised of a fast CPU (or multi-CPU machine), RAM, and disk storage is required for this functionality. The primary selection criteria for the hardware is that it should be supported by the portal software selected (the most critical software component).

The following software components will be required: a web server (e.g. Apache, Netscape Enterprise, MS IIS), a database with accompanying report generation software (e.g. PostgreSQL, Oracle, MS SQL Server), firewall software (e.g. IPFW, ipfilter, Firewall-1), a web forum (e.g. UltimateBulletinBoard, WWWboard), and a LISTSERV (e.g. MailMan, majordomo, LISTPROC), and portal software (e.g. Plumtree, Excalibur, Knowledge Center). Note that the web server, database, and web spider are also part of the compliance component described above. The same software can be used for both purposes.

### **Storage Component**

The storage component represents the "back-end" functionality serving the needs of the front-end components rather than the end-users directly. The data it contains consists of the documents required to be published by participants in accordance with the LSN Rule, and accompanying required information. Access will be interactive (e.g. when a home site front-end user requests a particular document, the home site front end will fetch it from the participant's repository). Access will also be by batch (e.g. the portal will fetch all materials on the web site, index them, and retain only the references to the documents for subsequent presentation in response to end-user queries). Each participant will assemble, prepare, and publish their own collections of documents on a WWW server. Participants will make their documentary collections available on a web server located at a site of their choosing and attached to the Internet. Participants are free to establish their own web server, collaborate on a community web server, procure commercial web service, or employ any other provisioning method they choose.

The compliance component and the presentation component will access these collections as WWW clients and perform the necessary operations routinely through participant sites.

For obtaining network usage statistics and performing monitoring activities, the compliance component will also require SNMP (Simple Network Management Protocol) access to participant web servers and network interface equipment.

Documents are to be converted to a format that includes an image representation (TIFF/CCITT or TIFF/JPEG), a searchable text file, and a bibliographic header containing metadata about the document. In many cases, this will require scanning and OCR conversion of a paper document. However, if a document exists in electronic format, it may be preferred to perform a more accurate conversion with appropriate software.

The LSNA may allow participants to provide their documentary collections in alternative page-representation formats such as PDF and proprietary word processor formats like Microsoft Word. This will depend on whether the data retrieval software selected for the front-end components is capable of indexing, searching, and otherwise processing these formats. The requirement to provide a bibliographic header for each document will remain regardless of the documents' formats. The bibliographic header is subject to the same retrieval requirements as the source document, e.g. provided as a searchable text file by the web server, as HTTP headers, or from within a database.

Document preparation is potentially the most labor-intensive and costly aspect of building the LSN, due to the large number of documents included. Therefore, the burden on a participant is more closely correlated to the number of documents they must prepare than any other factor.

Under this alternative, participants will place their documents on the web server of their choice through whatever file transfer mechanism is supported by the web server. This web server must be connected full-time to the Internet through a communications circuit of adequate speed (to be determined by the LSNA) and have a unique IP address and domain name. The domain name and root URL for the documentary collection, and a list of documents, must be provided to the LSNA.

For consistency in retrieval by the front-end components, participants may be required to follow a standard format in layout of the web pages that provide access to the documents themselves and accompanying bibliographic header information. Note that many web servers provide a standard way to publish meta information on web-served documents (e.g. by including this information in a file of the same name as the source document in a meta sub-directory). Use of this function may be required by the data retrieval elements of the front-end components.

Access from the portal/presentation site to the participant sites must be performed by software with fixed expectations of participant site structure and content. This will require that participants coordinate their site design and operations with the LSNA, which is expected to be a significant on-going operational requirement.

It is difficult to determine the exact hardware and software components due to the possibility of collaboration and the differences in the size of the documentary collections of the participants. Foreseeable alternatives for setting up a web server include a dedicated resource at the participant's site, sharing a server with other participants or non-LSN-related web sites, "co-location" of a participant-owned machine at an IPP (Internet Presence Provider) or outsourcing

the entire site to an IPP. Each of these alternatives have a wide range of cost, convenience, assuredness, and administrative issues associated with them.

If a participant adopts a strategy of implementing a dedicated web server, the size of this machine will, again, depend on the size of the document collection the participant is required to make available.

Participants with an extremely small document collection will probably choose to lease web space on an IPP machine or "piggy-back" on another participant's site rather than implement their own web server. The cost of this facility depends on the amount of data published, the bandwidth the site requires, and other metrics. Typical costs for web sites that are appropriate for small participants range from free (of incremental cost over maintaining a basic Internet-access capability) to several hundreds of dollars per month.

For those who choose to implement their own dedicated resource, a fairly modest machine may be fully satisfactory. An example of this would be an i386 architecture "PC" (e.g. 166MHz Pentium, 128MB RAM, 4GB disk) running an open-source Unix-like operating system (FreeBSD or Linux) and the open-source Apache web server. The total cost (hardware and software) of such a machine at current (4<sup>th</sup> quarter 1999) market prices is under \$1,000, and it would accommodate as many as 10,000 documents (at an estimated 250KB per document). Note that operational costs may not be so trivial, especially the disaster recovery aspects (regular backups with off-site storage), and data communications costs. However, resources for these requirements may already exist, and participants who choose to share a web server may be able to equitably spread these costs among themselves.

Participants with larger document collections will, naturally, require a more powerful computer system, and operational costs will scale as well. Due to the considerable resources of these participants and the likelihood of their already possessing significant computer system infrastructure, no attempt has been made to develop a cost for these facilities.

## **Cost**

The aggregate cost (to both the LSNA and the participants) of this design is lower than CPIC Alternatives #3. It has the additional characteristic that each participant's implementation costs are a function of the size of their documentary collections. Because participants are free to select technologies of their choice with their own cost factors included in the selection process, this will probably result in cost savings to them. Due to the increased demand on the participant sites for operational readiness and performance, it is likely that some of the life-cycle cost components will be higher, specifically, maintenance and communications. However, it is estimated that the overall cost to the participants (excluding their individual document preparation costs) will vary by about ten percent (10%) among the three alternatives.

The cost to the NRC to build and maintain the compliance and presentation component is significantly lower than under Alternative Three (known as TWG Alternative Five).

### **1.9.3 CPIC Alternative #3 - Portal with Comprehensive Interface with Participant Files Copied to a Central Store Operated by LSN**

As noted in Section 1.9.2, this design alternative differs from CPIC Alternative #2 primarily in the details of how and where LSN materials are stored.

This alternative has been named the "consolidated storage" alternative by the members of the TWG. This terminology focuses on a key characteristic of the design - that each participant assemble, prepare, and publish their own collections of documents on a WWW server that they control and place at a site of their choosing, and that their information is replicated at a single central storage facility maintained by the LSNA. This approach is essentially identical to the design originally conceived in the early conceptual development of the LSN, with additional storage capability included, effectively used as a 100% cache of participant documentary materials. This functionality allows improved performance, reliability, ease of implementation, and lessens the operational burden on participants.

The following sections will describe the details of its components, how they "fit" together, the hardware and software used in the design, and an approximation of the component life-cycle costs. The design of the first and second components are not materially affected by the alternative selected for the third component, although there are differences in the details of implementation and operation.

### **Compliance Component**

This component is used by a small set of LSNA users who require specific information at specific times to fulfill data assuredness requirements. It provides a mechanism to scrutinize the effectiveness of the LSN, to ensure that it is functioning as intended to the end-user base. Many functions of this subsystem will be accessed through a WWW browser using standard HTTP/HTML mechanisms bolstered by CGI programs. It includes a data retrieval element that routinely "roves" participant sites, fetching participant data (documents, statistics, and other) and storing this data pending processing. The primary method of following the operation and evolution of the LSN is through reporting mechanisms and this component provides the necessary reports on LSN functionality to enable the LSNA to ascertain participant compliance and aid in determining whether remedial action is required. Reports will be generated automatically by the system on a periodic basis, when exceptional conditions arise, and on-demand. The full array of required reports is yet to be determined. Minimally required are: a) listings of changes in participant document collections, i.e. additions, deletions, and modifications, and, b) the "health" of the LSN, component and sub-component uptime and performance data (e.g. web server hits, average response times, number of users, etc.). Historical trends will be maintained both in HTML tables and graphically. Collected data is stored and processed, and the results of that processing are likewise stored to build a file for trend analysis. The database will be a network-capable SQL relational database that will provide structured data to both front ends, i.e. the compliance and presentation components. Both file system storage and database storage will be accommodated. It also gathers low-level statistics on network operation for trend and throughput analysis.

This element provides a level of assuredness that the systems housing the LSN are functioning as required, specifically validating security mechanisms, network monitoring to evaluate hardware and software outages or sub-optimal operation, and physical plant including environmental and power conditioning. This sub-element provides appropriate and implements disaster recovery mechanisms, e.g. a backup/restore capability.

A single computer system of the workstation class is adequate for this functionality. The security sub-element mandates that the system be separate from and more restricted than the computer system (described below) that provides general access. The system should be equipped with the standard components, a graphical display, and a device appropriate for backup. The following software components will be required: a web server (e.g. Apache, Netscape Enterprise, MS IIS), a database with accompanying report generation software (e.g. PostgreSQL, Oracle, MS SQL Server), firewall software (e.g. IPFW, ipfilter, Firewall-1), network monitoring and management software (e.g. Big Brother, SunNet Manager, HP OpenView), and a web spider (e.g. MoMspider, BRS/Search, Fulcrum Search Server). Note that the web server, database, and web spider are also part of the presentation component described below). The same software can be used for both purposes. In addition, it is anticipated that this component will require some custom software, scripts and CGI's rather than full-blown applications.

### **Presentation Component**

This component is a "front-end" with a large set of users who require access to a wide range of information at arbitrary times. It is intended to fulfill the requirement to provide search and retrieval and information delivery to the parties through WWW technology. The intended user base includes all participants and potential participants, the LSNA and his designees, the press, and the general public.

It is a WWW presentation interface with additional sub-components that consist of:

- Introductory and overview documentation.
- Training / tutorial materials on how to use the site to obtain LSN-related information, and the other aspects of the site, and how to submit to the docket.
- Portal software that allows user customization of user interfaces and user document search and access strategies.
- A search facility that allows LSN-wide searching of participant materials, including individual user custom searching strategies.
- Publication of statistical information on LSN participant sites, including site content and performance.
- A web-based interactive forum in which interested parties can discuss or exchange information regarding LSN matters.
- Help-desk assistance (with escalation) for participants and public users.
- A LISTSERV (e-mail list manager) to allow participants to easily send electronic mail to all interested parties. A number of mailing lists will be created as needed for discussion of specific subjects, including a list with the e-mail addresses of all participants for notification purposes. The LISTSERV software will allow each participant to manage their own subscriptions to interest lists and archive

messages to the lists. The LSN is not intended to provide a public LISTSERV function.

- Aggregation and publication of overall LSN access and usage statistics, e.g., number of hits.

Web browsers will be the predominant access method to this component. It is anticipated that this will be the sole access method for the majority of users. Browsers will be used to gain access to general information, participant documentary collections, and to discussion forums.

A single server class computer system comprised of a fast CPU (or multi-CPU machine), RAM, and disk storage is required for this functionality. The primary selection criteria for the hardware is that it should be supported by the portal software selected (the most critical software component).

The following software components will be required: a web server (e.g. Apache, Netscape Enterprise, MS IIS), a database with accompanying report generation software (e.g. PostgreSQL, Oracle, MS SQL Server), firewall software (e.g. IPFW, ipfilter, Firewall-1), a web forum (e.g. UltimateBulletinBoard, WWWboard), and a LISTSERV (e.g. MailMan, majordomo, LISTPROC), and portal software (e.g. Plumtree, Excalibur, Knowledge Center). Note that the web server, database, and web spider are also part of the compliance component described above. The same software can be used for both purposes.

### **Storage Component**

This alternative design differs from CPIC Alternative #2 primarily by the introduction of a consolidated storage server that operates independently of the data stored at the participant sites. The storage component represents the "back-end" functionality serving the needs of the front-end components rather than the end-users directly. The data it contains consists of the documents required to be published by participants in accordance with the LSN Rule, and accompanying required information. Access will be interactive (e.g. when a home site front-end user requests a particular document, the home site front end will fetch it from the participant's repository). Access will also be by batch (e.g. the portal will fetch all materials on the web site, index them, and retain only the references to the documents for subsequent presentation in response to end-user queries). Each participant will assemble, prepare, and publish their own collections of documents on a WWW server. Participants will make their documentary collections available on a web server located at a site of their choosing and attached to the Internet. Participants are free to establish their own web server, collaborate on a community web server, procure commercial web service, or employ any other provisioning method they choose.

However, in this design, a server intended to aggregate all LSN data will be implemented in close network proximity to the compliance and presentation components. The storage server will load itself from the participant web sites and subsequently be accessed by the compliance and presentation components. It will act as a 100% cache for participant sites, loading itself from their sites and then supplying the front-end components with data. This will allow a less rigorous design in participant sites, because they will not be required to provide high levels of performance and operational characteristics. The consolidated server can access participant

sites at a slow rate of speed and at convenient times (for the participants), and then provide this data at high rates of speed and at all times to the presentation and compliance components.

The compliance component and the presentation component will access the consolidated storage component directly through a file system rather than through a web server for greater efficiency in data transfer.

For obtaining network usage statistics and performing monitoring activities, the compliance component will also require SNMP (Simple Network Management Protocol) access to participant web servers and network interface equipment.

Documents are to be converted to a format that includes an image representation (TIFF/CCITT or TIFF/JPEG), a searchable text file, and a bibliographic header containing metadata about the document. In many cases, this will require scanning and OCR conversion of a paper document. However, if a document exists in electronic format, it may be preferred to perform a more accurate conversion with appropriate software.

The LSNA may allow participants to provide their documentary collections in alternative page-representation formats such as PDF and proprietary word processor formats like Microsoft Word. This will depend on whether the data retrieval software selected for the front-end components is capable of indexing, searching, and otherwise processing these formats. The requirement to provide a bibliographic header for each document will remain regardless of the documents' formats. The bibliographic header is subject to the same retrieval requirements as the source document, e.g. provided as a searchable text file by the web server, as HTTP headers, or from within a database.

Document preparation is potentially the most labor-intensive and costly aspect of building the LSN, due to the large number of documents included. Therefore, the burden on a participant is more closely correlated to the number of documents they must prepare than any other factor.

Under this alternative, participants will place their documents on the web server of their choice through whatever file transfer mechanism is supported by the web server. This web server must be connected full-time to the Internet through a communications circuit of adequate speed (to be determined by the LSNA) and have a unique IP address and domain name. The domain name and root URL for the documentary collection, and a list of documents, must be provided to the LSNA.

For consistency in retrieval by the front-end components, participants may be required to follow a standard format in layout of the web pages that provide access to the documents themselves and accompanying bibliographic header information. Note that many web servers provide a standard way to publish meta information on web-served documents (e.g. by including this information in a file of the same name as the source document in a meta sub-directory). Use of this function may be required by the data retrieval elements of the front-end components.

Access from the portal/presentation site to the participant sites must be performed by software with fixed expectations of participant site structure and content. This will require that participants coordinate their site design and operations with the LSNA, which is expected to be a significant on-going operational requirement.

It is difficult to determine the exact hardware and software components due to the possibility of collaboration and the differences in the size of the documentary collections of the participants. Foreseeable alternatives for setting up a web server include a dedicated resource at the participant's site, sharing a server with other participants or non-LSN-related web sites, "co-location" of a participant-owned machine at an IPP (Internet Presence Provider) or outsourcing the entire site to an IPP. Each of these alternatives have a wide range of cost, convenience, assuredness, and administrative issues associated with them.

If a participant adopts a strategy of provisioning a dedicated web server, the size of this machine will, again, depend on the size of the document collection the participant is required to make available.

Participants with an extremely small document collection will probably choose to lease web space on an IPP machine or "piggy-back" on another participant's site rather than implement their own web server. The cost of this facility depends on the amount of data published, the bandwidth the site requires, and other metrics. Typical costs for web sites that are appropriate for small participants range from free (of incremental cost over maintaining a basic Internet-access capability) to several hundreds of dollars per month.

For those who choose to implement their own dedicated resource, a fairly modest machine may be fully satisfactory. An example of this would be an i386 architecture "PC" (e.g. 166MHz Pentium, 128MB RAM, 4GB disk) running an open-source Unix-like operating system (FreeBSD or Linux) and the open-source Apache web server. The total cost (hardware and software) of such a machine at current (4<sup>th</sup> quarter 1999) market prices is under \$1,000, and it would accommodate as many as 10,000 documents (at an estimated 250KB per document). Note that operational costs may not be so trivial, especially the disaster recovery aspects (regular backups with off-site storage), and data communications costs. However, resources for these requirements may already exist, and participants who choose to share a web server may be able to equitably spread these costs among themselves.

Participants with larger document collections will, naturally, require a more powerful computer system, and operational costs will scale as well. Due to the considerable resources of these participants and the likelihood of their already possessing significant computer system infrastructure, no attempt has been made to develop a cost for these facilities.

The storage server will be a high-capacity, high-performance computer capable of housing a fairly large amount of data (approximately double the size of the entire documentary collections of all the participants), estimated to be in the four terabyte range. An example of a system in this class is a Sun Microsystems model 10000 in an appropriate configuration. The software to provide the file sharing capability is generally included with the base system and the software mentioned in the description of the portal will be used to populate the storage server so there is no additional software component.

## **Cost**

The aggregate cost (to both the LSNA and the participants) of this design is significantly higher than Alternatives Two (known as TWG Alternative Three). This design has the characteristic

that each participant's implementation costs are a function of the size of their documentary collections. Because participants are free to select technologies of their choice with their own cost factors included in the selection process, this will probably result in cost savings on their behalf. Due to the decreased demand on the participant sites for operational readiness and performance, it is likely that the overall life-cycle cost components will be lower, specifically in site maintenance and communications costs. However, it is estimated that the overall cost to the participants (excluding their individual document preparation costs) will only vary by about ten percent (10%) among the three alternatives.

The cost to the NRC to build and maintain its portion of the LSN is estimated to be significantly higher than under CPIC Alternative #2 due to the need to acquire and maintain the storage server. Such systems are well within the capabilities of modern computing systems but the raw storage needed, and the processing power to manage that much data can be costly.