3. SOLUTION FOR LSN IMPLEMENTATION

3.1 Technical Focus

The technical focus of the LSN is to provide information through the Internet to the entire community of potential parties, interested governmental participants, and parties to the high-level waste (HLW) repository proceedings in an as-simple-to-build-and-maintain a manner as practicable. The LSN web portal provides a central point of access to the general public and LSN participants to all LSN information, including such additional elements as:

- A centralized search facility that, when queried, will survey bibliographic header data and content on participant sites and return references with Hypertext Transfer Protocol (HTTP) links to all matching documents on all participant sites;

- Publication of statistical information about participant LSN websites, including site activity and performance; and

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

The LSN as a web portal is not a central repository, but the central source for discovery information for the HLW licensing proceeding. Although documents from participant sites are not stored on the LSN web portal, the content and context are accessible via its search engine. The LSN web portal stores pointers or HTTP links to the actual documents on an individual participant's web server. When users find the document they seek and request it, the document is served up from the participant's LSN website to the requester. The LSN web portal contains only a limited number of active components and will remain relatively static with few manual updates.

The archiving of the data (documents) is the responsibility of the participant. The participant also is responsible for making the data available to the LSN as described in 10 C.F.R. Part 2, Subpart J.

3.2 LSN Design Overview

The following subsystems represent major capabilities of the system:

- Participant Document Fetch
- Participant Site Audit
- LSN Search
- LSN Administration
- LSN General Web Content (includes Participant Representative Login)

3.3 Hardware and Software Configuration

The LSN web portal is assembled from standard hardware, operating systems, Commercial-Off-The-Shelf (COTS) products, and reusable components with some custom software. The components are glued together with lightweight scripting and hypertext. The website has easy-to-use web pages that deliver content to participants, the public, and the LSNA and staff.

The LSN web portal presents a home page, along with other general web content. Administrative functions, such as controlling an access list of the priority participants, are available only after the program is installed on a computer and requires a CD in the tray to operate.

The primary search and retrieval software underlying the LSN site is the Autonomy™ software package. It provides the indexing of participant LSN websites plus intelligent search and retrieval services. Autonomy performs searches based on the content and context of the searcher's requirements. Searches based on context, rather than key words or phrases, return only those documents that respond to the searcher's specific requirements instead of all documents that contain a specific key word or words that may be in a document, but are not relevant in the context of the searcher's subject. A custom application was developed to determine during the crawl of each participant's repository if any documents were added, changed, or deleted.

Document auditing metrics are consolidated into a Structured Query Language (SQL) server database and uniformly presented to the LSNA through the LSN Administration System. The LSN Administration System provides network performance and connectivity monitoring. Autonomy log files provide statistics on bibliographic headers, documents, and image files add/change/delete transactions as it "crawls" each participant's website. A custom application monitors participant site availability.

Access to the LSN for priority users is managed through a SQL server database and Active Server Page (ASP) script. Participants must have a user ID and password to login as a priority user. The LSN web portal monitors the servers for load balancing abilities. During times of heavy usage, the public may lose the ability to perform searches, allowing only participant users who are logged in to be able to perform searches.

3.4 Major LSN Repositories

There are five primary stores of information. Significant amounts of LSN data are stored outside of a relational database. Autonomy, for example, manages its own store of information, optimized for free text searching. Some of these repositories are "work areas" where data is staged for processing, not for end-user searching. The major LSN data repositories are:

- Autonomy Staged Content Data – an Autonomy work area where document content is initially staged as it is retrieved from participant sites.

- Autonomy Production Content Data – the primary searchable repository. It contains both searchable content and bibliographic header information merged into one Autonomy record. It also contains older records, deleted by participants.

- SQL Header Database – a database work area where bibliographic headers are stored. It contains current bibliographic headers plus older changed and deleted bibliographic headers.

- SQL Audit Data – an audit database of add/change/delete transactions for both documents and bibliographic headers.

- SQL Participant Site and Representative Database – contains information on participant sites, facilitating indexing and retrieval of participant documents.

3.5 Participant Document Fetch

The Participant Document Fetch subsystem creates a searchable LSN repository of unstructured document content and structured bibliographic headers.

The first phase of the Fetch process uses a working area or staging repository to prepare the data. The Autonomy HTTP Fetch retrieves the textual content of all documents and stores them in the Autonomy Staged Content Data. The Header Importer, a custom Microsoft .NET component, fetches all bibliographic headers, parses them, and stores them in the SQL Header Database staging area.

In the second phase of the Fetch process, the Header-Content Merger, another custom .NET component, pulls each staged bibliographic header from the SQL Header Database and matches it by Uniform Resource Locator (URL) to the corresponding staged document content from Autonomy. The Header-Content Merger component then combines both into one Autonomy record and writes it to the Autonomy Production Content Data repository. Bibliographic header data is added as custom fields to the Autonomy record, with any multiple occurring fields concatenated together.

Users will search bibliographic headers and content in the Autonomy Production Content Data repository. Combining bibliographic headers and content into one record allows users to write advanced Boolean searches that combine both content keywords and bibliographic header terms into one query. The SQL Header Data in the staging area is not searched, but is used when displaying a header to the user.

3.6 Participant Site Audit

As the fetching process runs, it also provides information needed to detect added, changed, and deleted documents. The SQL Header Database is a set of relational tables containing all bibliographic header data (“Metadata”) retrieved from participant sites. The single-value fields (such as Title, Participant Accession Number, and LSN Accession Number) are in one central table (Header); the multi-valued fields are in subsidiary tables (one each for Authors and

Organizations, Addresses, and Related Records, for example). This table is populated by the Header Importer component. During the spidering process, a custom application detects any byte changes in a document that has been placed in a participant's document repository.

Fetch Auditor examines each entry in the Autonomy Fetch Results log and attempts to match it to a corresponding Stages and Production record. This locates changed and deleted documents. The Fetch Auditor also identifies deleted bibliographic headers. The Fetch Auditor is driven by the content of the log file created by the Autonomy HTTP Fetch. The Fetch Auditor is built using the Autonomy Search Agent, which provides a Component Object Model (COM) interface (API, or application program interface) to the Autonomy data, and the ActiveX Data Objects (ADO) Fetch Audit Data, which provides a COM interface to the SQL Server database of audit events.

Specifically, it will detect any byte changes in a document that has been placed in a participant's document library. Fetch Auditor examines each entry in the Autonomy Fetch Results log and attempts to match it to a corresponding Staged and Production record. This locates changed and deleted documents. The Fetch Auditor also identifies deleted bibliographic headers.

3.7 LSN Search

The Autonomy Production Content Data contains a complete set of document content and bibliographic header fields indexed for searching.

The Basic Search is an ASP page containing general LSN content (e.g., navigation links), search help, and an HTML form. The Basic Search Form is an HTML format that has basic search input fields. The Advanced Search is the equivalent ASP page for an advanced search. It contains the Advanced Search form, which has an input field for each searchable item (e.g., bibliographic header fields, participant sites, and content search terms).

The user fills out the appropriate form and presses the "Submit" button. This sends the contents of the form to the Search Summary ASP page. This page contains the Autonomy Search Agent component and some subroutines to format the search and run the Search Agent. The component searches the Autonomy Production Content Data and assembles the results. Then the Search Summary ASP page generates the Search Summary HTM page. The LSN retrieves each search field from the component and wraps HTML tags around the fields to format the output. The web server delivers this generated HTML, Search Summary HTM, to the user's browser.

Both search forms submit their search criteria to the Search Summary page. The Search Summary is a server-side page that contains code to formulate and execute the search. It scripts an Autonomy-supplied component to execute the search and format the results for display in HTML. The Autonomy Search Agent is the scripted component that provides Autonomy search services. The Search Summary HTM is a client side page generated by the corresponding ASP page.

The user may click on a link on the summary page to see more details or go directly to the document on the participant site. The LSN does not directly deliver up documents or images. It merely provides a page that will hyperlink the user to the documents or images located on the participant sites.

3.8 LSN Administration

Administration of the LSN is accomplished through several integrated subsystems.

These subsystems include:

- Custom designed server components which provide accumulated processing information
  - LSN Spider
  - LSN Auditor
  - LSN Site Poller
- Custom designed Microsoft Windows-based Administration Interface Application which allows LSN Administrators to maintain the website content, manage organizations and users, view system status, view system reports, and manage document modifications. This application is protected by encrypted user identification and passwords, along with a verification compact disc.

Direct system maintenance of the servers and server software is provided through a secure frame-relay connection, which allows for the updating of software, the application of security patches, and retrieval of archival information.

3.9 LSN General Web Content

Participant Login is an ASP page where participant representatives enter their user IDs and passwords. The password is not displayed on the screen as it is typed and is sent to the server in a secure (encrypted) manner. LSN Home is the standard home page. Participants are automatically redirected here after successfully logging on. Login Help provides additional help and an LSN Webmaster contact point for logon or password problems.

This page performs straightforward ASP Form and ADO object scripting. The user enters his or her user ID and password. The script constructs the SQL query using the form fields to build an appropriate WHERE clause (for example: WHERE UserID='xxx' and Password='yyy'), connects to the database, executes the query, and receives an ADP Recordset back from SQL.

If there are no rows in the Recordset, either the ID or the Password or both were incorrect. The script generates an appropriate error and counts the failed login attempt. If there is exactly one row in the Recordset, the login was successful. The script stores the participant type (role) and participant ID in the ADO Session object.

Optionally, if the failed login attempt exceeds a threshold, the script disables the Participant's account and generates an e-mail to the administrator.

If the user is changing his or her password, the page checks for a minimally-acceptable password, prompts the user to re-enter the password, and records it in the database.

LSN Home is the home page. It includes in its HTML two other page fragments, search and navigation. Simple Search is a page fragment that includes a simple search form. As the primary purpose of the LSN is to provide a search capability, a simple search is commonly available on most LSN pages. Advanced Search is a separate page navigated to by a single click.

Navigation Options is another page fragment that is included on most pages. It is a dynamically constructed "navigation bar" containing hyperlinks to all common LSN pages. The Navigation Options page fragment uses information in the ASP Session to dynamically configure itself for each user type. For example, guests will not be able to save searches. ASP Session is a built-in ASP object that contains the participant type for the logged on user. The current types are "participant" and "guest" (anyone not logged on).

Fanning out from Navigation Options is the set of general content pages: About LSN, Calendar of Events, LSN Participants, etc. This is straightforward web content, not requiring any object-oriented modeling.

3.10 Secure Hosting Solution

The hosting is provided by AT&T Internet Data Centers and Global Network. Participants and the public will experience consistently fast access to the central LSN site. Reliability is designed into the network infrastructure through the use of proven network components and backup or fail-over hardware, and by ensuring that there are no single points of failure. Multiple paths to redundant facilities assure the highest degree on online availability possible.

AT&T staffs the Network Operating Center for the Internet Data Center and Global Internet Protocol (IP) Backbone with technicians that monitor the IP networks and Internet Data Center internal networks for overall network and bandwidth optimization, as well as bandwidth utilization, 24 hours a day, seven days a week (24x7). Technical support is also available 24x7 for the LSNA to report web server or network connectivity problems.