

3. SOLUTION FOR LSN IMPLEMENTATION

3.1 Technical Focus

The technical focus of the LSN is to provide information through the web to the entire community of potential parties, interested governmental participants, and parties to the high-level radioactive waste repository proceedings in an as-simple-to-build-and-maintain a manner as practicable. The LSN web portal will provide 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 high-level waste repository 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 will contain 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 will be assembled from standard hardware, operating systems, Commercial-Off-The-Shelf (COTS) products, and reusable components with a minimum of custom software. The components will be glued together with lightweight scripting and hypertext. The website will have 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 and provides additional administrative functions, such as controlling an access list of the priority participants.

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.

Although auditing will be done with a mix of products, all document auditing metrics will be consolidated into a Structured Query Language (SQL) server database and uniformly presented to the LSNA through limited access hypertext pages. WebTrends and WhatsUp Gold products will provide 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 at the byte level.

The operating system is Microsoft Windows 2000. The hardware servers are Compaq Proliants, which will run the Windows 2000 operating system, web server software, and Microsoft SQL Server 2000 database.

Access to the LSN for priority users will be managed through a SQL server database and Active Server Page (ASP) script. Participants will have a user ID and password to login on the LSN web portal. When a participant logs in they will be switched to a specific path into the network that will provide processing on a dedicated server. All other users will remain on the public server. The LSN web portal will monitor both servers for load balancing abilities. If the public server is not being utilized and there is heavy traffic on the participant server, then some participant searches may be performed against the public server. This feature would not be apparent to the user.

3.4 Major LSN Repositories

There are six 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.
- **Site Server Log Files and Availability Audit Statistics** – are maintained within two COTS products, WebTrends and WhatsUp Gold. WebTrends retrieves common log files and builds reports for each participant site. WhatsUp Gold monitors participant site availability and response time.
- **SQL Participant Site and Representative Database** – contains information on participants, especially IDs and passwords.

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 Visual Basic (VB) 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 VB 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 remains available for future uses if needed.

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.

SQL Header Audit Data contains a record of the bibliographic header audit of participant sites. Events such as added, changed, and deleted bibliographic headers are detected and recorded here by the Header Importer component. SQL Document Audit Data contains a similar log of added, changed, and deleted documents detected by the Header-Content Merger and the Fetch Auditor.

WebTrends retrieves common log files and builds reports for each participant site. WhatsUp Gold monitors participant site availability and response time. 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. Automated extracts imported from participants' web logs provide document usage information. 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 Common Log File Format

Server logs are files produced by any computer that stores and transmits documents to other computers on a network every time a document is requested. It resides on the computer server and has entries containing basic information about each individual HTTP request and its result. The standard format for LSN web server logs is the National Center for Supercomputing Applications' (NCSA) common log file format. When installing a web server or working with an Internet Service Provider (ISP), the system administrator specifies in which format to save the log file. While most sites use one of the two dominant server applications -- either Microsoft's Internet Information Server (IIS) or Apache -- there may be sites that use another server application. When setting up a participant's web server, the system administrator should turn on logging and choose NCSA Common Log Format.

The LSN will use secured File Transfer Protocol (FTP) to retrieve a copy of the log file from the dedicated area on the web server. After the log file has been copied to the LSN audit server, the data will be processed to produce web-based statistics using WebTrends software.

The format of the common log file should be composed of the following space-delimited fields, with a single dash placeholder for each absent value:

Host Ident Authuser Date Request Status Bytes

| | |
|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Host | Domain name of the client or its IP number if the name is not available. For example, the domain name might be Isnnet.gov or, as in the sample log file below, the IP number 198.22.152.217 is provided in lieu of a domain name. |
| Ident | If the identify check is enabled and the client machine runs Ident, then the identify information (the user's name) is reported. In the example below, the client has not turned on Ident so there is no value and Ident is represented by a hyphen (-). |
| Authuser | If the request is for a password protected document, this is the userid used in the request. In the example below, the client has not turned Authuser on so there is no value and Authuser is represented by a hyphen (-). |
| Date | [day/month/year:hour:minute:second zone] The day is two digits. In the sample below it is the value 26. The month field reserves three letters. In the sample below it is the value Apr. (i.e., April). The year field reserves four digits. In the sample below it is the value 2001. The hour field reserves two digits. In the sample below it is the value 13 (i.e., 1 p.m.). The minute field reserves two digits. In the sample below it is the value 59. The second field reserves two digits. In the sample below it is the range from 42 to 50. The zone is ('+' '-') four digits. In the sample below it is the value -0500. |
| Request | Identifies and captures, in double quotes, the command requested from the client and the file name that is being requested. Some of the various GETs are seen below as the user navigates around the web site looking at various pages. |
| Status | A three-digit code returned to the client. The status codes have universal meaning and are used by WebTrends to generate web site statistics. For example, status code 200 means that the file was successfully retrieved. |
| Bytes | Bytes is the number of bytes in the object returned to the client, not including any headers. In the first line of the sample below, the home.asp returned 0 bytes, and in the second line, the logo.gif returned 140. |

```
198.22.152.217 -- [26/Apr/2001:13:59:42 -0500] "GET /home.asp HTTP/1.1" 200 0
198.22.152.217 -- [26/Apr/2001:13:59:42 -0500] "GET /images/logo.gif HTTP/1.1" 304 140
198.22.152.217 -- [26/Apr/2001:13:59:42 -0500] "GET /images/logo_background.gif HTTP/1.1" 304 140
198.22.152.217 -- [26/Apr/2001:13:59:42 -0500] "GET /images/spacer.gif HTTP/1.1" 304 140
198.22.152.217 -- [26/Apr/2001:13:59:42 -0500] "GET /images/preview.gif HTTP/1.1" 304 141
198.22.152.217 -- [26/Apr/2001:13:59:42 -0500] "GET /images/nav_dot.gif HTTP/1.1" 304 140
198.22.152.217 -- [26/Apr/2001:13:59:42 -0500] "GET /images/bullet.gif HTTP/1.1" 304 140
198.22.152.217 -- [26/Apr/2001:13:59:42 -0500] "GET /images/hline.gif HTTP/1.1" 304 140
198.22.152.217 -- [26/Apr/2001:13:59:42 -0500] "GET /images/bullet_search.gif HTTP/1.1" 304 140
198.22.152.217 -- [26/Apr/2001:13:59:44 -0500] "GET /Docket.asp HTTP/1.1" 200 0
198.22.152.217 -- [26/Apr/2001:13:59:45 -0500] "GET /Home/Hearings.asp HTTP/1.1" 200 0
198.22.152.217 -- [26/Apr/2001:13:59:45 -0500] "GET /images/Construction.gif HTTP/1.1" 304 141
198.22.152.217 -- [26/Apr/2001:13:59:45 -0500] "GET /Home/Participants.asp HTTP/1.1" 200 0
198.22.152.217 -- [26/Apr/2001:13:59:46 -0500] "GET /Home/Operations.asp HTTP/1.1" 200 0
198.22.152.217 -- [26/Apr/2001:13:59:47 -0500] "GET /Motions.asp HTTP/1.1" 200 0
198.22.152.217 -- [26/Apr/2001:13:59:47 -0500] "GET /Help.asp HTTP/1.1" 200 0
198.22.152.217 -- [26/Apr/2001:13:59:48 -0500] "GET /Docket.asp HTTP/1.1" 200 0
198.22.152.217 -- [26/Apr/2001:13:59:49 -0500] "GET /Login.asp HTTP/1.1" 200 0
198.22.152.217 -- [26/Apr/2001:13:59:50 -0500] "GET /Home.asp HTTP/1.1" 200 0
```

3.8 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), several search hints, 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 VB Script 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 HTML page. VB Script 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 HTML, to the user's browser.

Both search forms submit their search criteria to the Search Summary page. The Search Summary ASP is an ASP server-side page that contains VB Script 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 HTML 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.

3.9 LSN Administration

Administration is an ASP page that acts as a menu of administrative functions, providing links to the four major administrative areas. This page and all other administration pages first check the ASP Session Object to ensure that the user has administrator rights.

ASP Session is a system component that sorts any user-specific information. There is one unique Session Object in memory for each current user of the system. The Session Object controls what each different type or class of users can see. When a participant logs on, the logon scripts store a participant type flag in the Session Object indicating public users, participant representative users, or NRC administrative users. The Administration page and all other relevant pages check this flag to prohibit access to any one page. If a user is not authorized to view a page, he or she is redirected away. Every page performs this check as its first action.

COTS products perform many of the functions for Administration. The administrative user navigates to them by clicking on a link on the Administration Page. Participant Performance Report is a set of web pages that display the information assembled by WhatsUp Gold. WhatsUp Gold collects statistics on each participant web site. Participant Log File Report is a set of web pages that displays common log file reports compiled by WebTrends. WebTrends builds reports from the participant common log files.

There are four Document and Header Audit reports. These reports are hypertext linked together, letting the administrator either point-and-click to drill down into details or select a report from the menu and fill in a search form. They are a list of administrative reports on participant sites' added, changed, and deleted bibliographic headers and documents. They are built in ASP using ADO to report from the SQL Fetch Audit Data and the SQL Header Data. The administrator may select either a report from a menu and fill in a form, or click on a report line to drill down to a more detailed report. The Summary Doc/Hdr Report lists a count of added/changed/deleted bibliographic headers and documents by each participant totaled across a selected date range. The Daily Doc/Hdr Report lists a count of added/changed/deleted documents and bibliographic headers by each participant by each day in the date range. The Doc Content Detail Report shows the document content for one or more added/changed/deleted document(s) selected by any or all of the following fields: LSN Accession Number, Participant Accession Number, Participant, and From-To date range. The Header Detail Report shows the bibliographic header content for one or more added/changed/deleted bibliographic header(s) from SQL Server Header Data.

The Administer Participants package provides a mechanism for maintaining the SQL Participant Site and Representative Database of participant organizations and participant representatives. The primary purpose is to let the LSN Administrators control the users who are authorized to log onto the participant-only servers, and to authenticate users by ID and password when they do log on.

LSN Maintenance provides supplemental contractor access to the operating system, database, and Autonomy software over a frame relay.

3.10 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. LSN Home is the standard home page. Participants are automatically redirected here after successfully logging on. Login Help provides additional help and an LSN Administrator contact point for logon or password problems.

This page performs straightforward ASP Form and ADO object scripting. The user enters his or her 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 should be 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 will be included on most pages. It will be 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, only administrators will see the "Administration" link, guests will not be able to save agents, etc. ASP Session is a built-in ASP object that will contain the participant type for the logged on user. The current types are "participant," "administrator," and "guest" (anyone not logged on).

Fanning out from Navigation Options is the set of general content pages: About LSN, FAQs, Privacy/Security Statement, Help, Searches, etc. This is straightforward web content, not requiring any object-oriented modeling. Administration is the administrator's menu. This and all other administrative pages will have security permissions set on them, requiring the administrator to have an NT account on the server. This will provide an extra measure of security for administrative functions. NT Security represents the Windows 2000 (NT-5) system-provided challenge/response security.

3.11 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 reporting web server or network connectivity problems.