

12. AUDITING SYSTEM

As part of its responsibilities under the LSN Rule, the LSNA will establish a system designed to monitor document integrity and system performance within the LSN program and to generate informational reports of such activity. The objectives of such an auditing system are to:

- Give participants reasonable assurance of the integrity of presented documentary material and of participant compliance with their responsibilities under the LSN Rule; and
- Obtain usable statistical data for informational and system management purposes.

12.1 Automated Auditing

The auditing system will be primarily an automated one, utilizing software that will scan participant LSN websites and gather data about website content and website activity. LSN staff will analyze the data gathered in this fashion and make determinations regarding participant compliance and system management issues. The types of data that the system will gather on each participant LSN website include:

- Count the number of records that have been added since the last monitoring crawl, including:
 - Number of bibliographic header records
 - Number of text files
 - Number of images
- Files that have been changed
- Files that have been removed
- Total number of documents to-date
- How many "hits" the website has recorded
- What response time the website has provided
- If the website has suffered any unscheduled downtime

The information gathered will assist the LSNA in measuring participant compliance and in assessing the effectiveness of LSN operation. Information obtained through monitoring will routinely be published on the central LSN site, after analysis and appropriate follow-up actions have been completed. For example, if a monitoring crawl of a participant's LSN website reveals that a document has been changed in some way, LSN staff will contact that participant to determine the nature of the change before publishing a notice of the change on the central LSN site. The LSN staff analysis will operate on an exception basis, investigating out-of-the-ordinary occurrences discovered by the automated monitoring crawl.

If a participant will be using an ISP to house their LSN text images and bibliographic headers, they must ensure the ISP will allow a crawler (spider, robot, etc.) to visit their LSN website. The participant must also ensure that the ISP will allow access to their log files for auditing purposes. To enable auditing, participants are requested to turn logging "on" and to store the log file in NCSA Common Log Format.

The participant should consider costs associated with cost-per-transfer, if applicable, for their LSN website. If the participant's ISP charges the participant on a transfer rate basis, the participant should contact the LSNA to make arrangements to schedule the crawling of their ISP hosted-site to minimize the financial impact on the participant. LSN staff will be coordinating with each participant to identify the crawler and schedule crawl time (bearing in mind server maintenance and ISP retrieval costs, etc.) at a later date.

12.2 Data Accuracy

Sampling of participant text data will be conducted by LSN staff by random sampling to confirm that document conversion accuracy approaches the recommended standards detailed in the Guidance Section on Optical Character Recognition (OCR) accuracy.