



# The Case for NSF— How Archiving in NSF Format Preserves Data Integrity

A Whitepaper by Grant Lindsay,  
Product Manager, Sherpa Software



**SHERPA**  
SOFTWARE

Sherpa Software  
456 Washington Rd. Ste 2  
Bridgeville, PA 15017  
(800) 255-5155  
[www.sherpasoftware.com](http://www.sherpasoftware.com)

## Introduction

The need for archiving electronic records, including electronic mail [email] continues to grow. Large volumes of electronic records, increased storage limitations, government regulations and legal implications have made the need to archive information a critical business issue. This trend will continue to grow over the next several years as electronic communication evolves and expands. This brings up several important considerations IT administrators should be mindful of. Between the time messages are captured and the time they are needed again for presentation or review, they will need to be stored. For Lotus Notes and Domino mail environments, this begs the question; is there a preferred storage format?

## Why Email Archiving is Important

Email archiving<sup>1</sup>, particularly compliance archiving<sup>2</sup>, is a growing business need. Email has moved from a convenient and innocuous means to expedite communication to a sensitive repository of critical business information. As the nature of email evolves, organizations are discovering an increased need to retain and protect their email assets, most commonly accomplished through email archiving.

Since collecting and maintaining these email archives involves additional costs, organizations may hesitate to implement a suitable archiving policy and procedure. Litigation brought by competitors, customers, or employees or fines levied by regulatory bodies may shock a company into utilizing and enforcing such a system, but only after significant and unnecessary expense.

On the other hand, many companies do not have a choice. Email retention standards are imposed upon them by external entities, often a federal government agency like the Securities and Exchange Commission [SEC]<sup>3</sup>. In the United States, the SEC requires publicly traded financial firms to archive business email for three years and to store messages in 'an easy-to-retrieve form' for two of those years<sup>4</sup>. Another U.S. example is the Federal Judiciary, which amended the Federal Rules of Civil Procedure in December 2006 and mandated that businesses be able to quickly produce all electronic content relating to a case<sup>5</sup>.

*"The reasons a company may opt to implement an email archiving solution include protection of mission critical data, record retention for regulatory requirements or litigation, and reducing production email server load."*

<sup>1</sup> See Appendix A for definition

<sup>2</sup> See Appendix A for definition

<sup>3</sup> See <http://sec.gov>

<sup>4</sup> Referenced from <http://www.financetech.com/featured/showArticle.jhtml?articleID=193004323>

<sup>5</sup> Referenced from [http://www.usdatavault.com/library/email\\_archiving\\_best\\_practices.pdf](http://www.usdatavault.com/library/email_archiving_best_practices.pdf)



Aside from industry or government regulations affecting organizations, it can be costly to ignore the need for email archiving. Lawsuits have been lost due to ineffective or missing email archiving policies and procedures.

Finally, an organization may choose to archive email messages for reasons related to business operations, for example, to protect contracts or other critical documents that live within the email system. Another reason to archive is to reduce the storage space and server strain placed on the overall IT environment. Regardless of the reason to archive, there are still many choices to be made before an archiving system is put into place.

## Data Storage Choices

Once an organization has decided to develop its email archiving and retention policy, it still faces many decisions. These include:

- How long messages should be kept?
- Which messages should be kept?
- What technologies exist to collect and store messages?
- What format should be used to store the collected messages?

The storage format choice should meet some basic requirements:

- **Scalability** – Will the storage method used be able to grow? Since archives usually grow faster than expected, several years worth of readily-accessible email messages can take up surprising amounts of space.
- **Searchability** – Can you find the needle in the haystack? Can you do it quickly? It is simplistic to say that there is no point in collecting data that will never be viewed again. However, it is fair to expect that an email archive will be searched at some point in its existence.
- **Portability** – How easy is it, or is it even possible, to move the data collected by one vendor's product to another's? Vendors and their products often come and go. An organization's ability to discover and produce search results can be severely limited if a proprietary data store is no longer available in "an easy-to-retrieve form." An organization may need to migrate all of its email archives to another system.
- **Security** – Are the archived messages protected from alteration or deletion? Once stored, a primary concern for compliance archiving is the assurance that the messages remain as they were when they were first stored.



- **Data Integrity** – Has the storage process altered the structure of the message? Is it important? Some courts have required that subpoenaed messages be presented in their customarily stored format, complete with attendant metadata. As a result, some experts are suggesting that "far from excluding this often critical information, the practitioner is well-advised to preserve metadata as a regular practice, particularly in connection with complex litigation."<sup>6</sup>

For IBM Lotus Notes and Domino shops, there is more than one option when deciding where to store archived email. This decision usually falls into either choosing "native" (i.e., NSF) storage or "converted" (e.g., relational, etc.) storage.

## Advantages of NSF (native) Format

Lotus Notes' native format, NSF<sup>7</sup>, offers strong characteristics in each of the areas mentioned above.

### *Elastic Capacity*

While each individual NSF file has a storage limit (currently 64 GB); any number of NSF files may be logically combined to create a data store of any size. Additionally, these NSF files may be added, moved or deleted as needed to manage the archive capacity.

### *Flexible Searching*

The Notes and Domino environment has a built-in mechanism to index and search data stored in NSF files. Once indexed, the searching of these data stores is quite fast, and no additional software is needed. However, since the NSF format has several published and documented application programming interfaces [APIs], there are many third-party applications that can make the process of searching NSF files even easier and faster.

### *Vendor Independence*

While it is true that the NSF format itself is owned and maintained by IBM, there is an extensive community of business partners and others who write applications using the documented NSF APIs. Thus, when using native format for Notes email messages, an organization is not locked into a format that only one vendor uses, but has the freedom of a format that many, many vendors use. This provides multiple choices when looking for a solution today and options for later, if the need arises to change email archiving products or vendors.

---

<sup>6</sup> Referenced from *The Increasing Importance of Metadata in Electronic Discovery*, by W. Lawrence Wescott II, published in the Richmond Journal of Law & Technology, Volume XIV, Issue 3

<sup>7</sup> See Appendix A for definition



Additionally, IBM currently supports the NSF format on a variety of operating systems, giving an organization flexibility to retain the existing email archiving data store even when moving from Microsoft Windows to Red Hat Linux, for example.

#### *Robust Security*

Included within the NSF structure is a mature and well-developed security model. This includes seven levels of access control and Public Key Infrastructure [PKI] encryption. Using these built-in facilities, access to a given NSF can be as locked down or as open as it needs to be, with fine-tuned, user-specific granularity. These features of the NSF format can ensure that no unauthorized viewing or alteration of archived messages can happen.

#### *Truly Unaltered Copies*

Since both the source and target data stores are the same, the messages can be copied with no alteration, keeping metadata intact. Then, during the later e-discovery presentation phase, the messages can be extracted in their native format or converted, at that time, to another format.

If, on the other hand, the format conversion happens when the message is stored, metadata may be lost. An organization's options for presentation formats may thus be limited and, depending on the situation, may be unacceptable.

## **Possible Advantages of Other (converted) Formats**

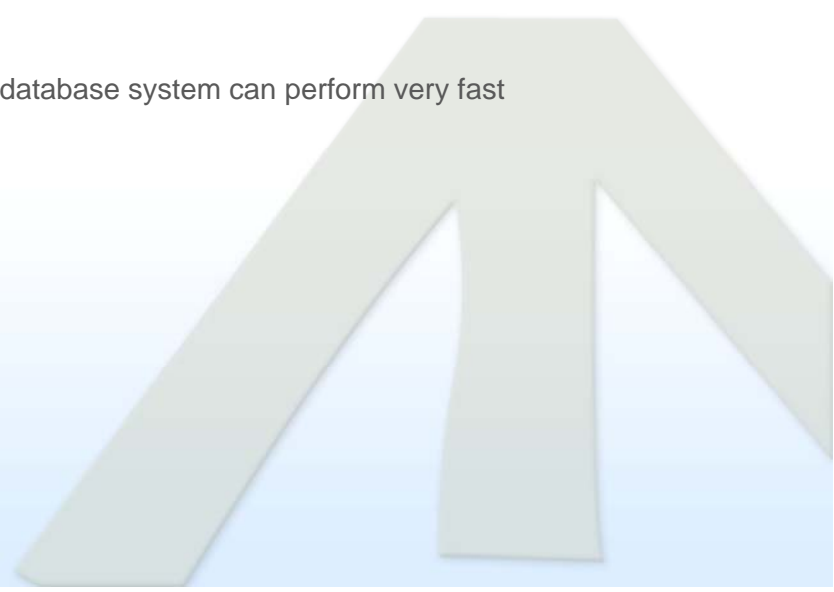
Alternatives to NSF for storing Notes data typically involve some type of relational database management system. There may be some advantages to choosing such a system, even though there will be an inherent data format change and potential loss of metadata.

#### *Scalable*

Tables in a relational database system can typically scale up to millions of rows and data normalization may improve storage consumption.

#### *Fast Searching*

With effective indexing, a relational database system can perform very fast searching.





## E-discovery Implications

Using native NSF format for Lotus Notes email messages need not impede e-discovery, in fact, it does just the opposite. With advance planning, messages can be stored in individual NSF files in an organized fashion that narrows the focus of an e-discovery request and shortens the turn-around time.

For example, messages may be stored in daily NSF files so, when a date range is applied to a search request, only a subset of the whole data store (a few NSF files) needs to be searched.

## Litigation Implications

Over the past decade, more than a few court cases have been decided based, in part, on the quality of the electronic evidence. Therefore, a prudent organization will make sure they have a secure repository of email stored in native format. Take Morgan Stanley and Merrill Lynch as examples:

*"In a May 2005 case in Florida, a jury ordered Morgan Stanley to pay more than \$1 billion in punitive and compensatory damages... The bank lost the case in part because it had failed to produce required e-mail and other documents in a timely fashion... In March 2006, Merrill Lynch was fined \$2.5 million by the SEC for lax archiving practices."<sup>8</sup>*

## Long-term View

Email archiving is an ongoing process, not a project. Organizations need to make an extended commitment in human capital and technology to have a successful email archiving process. This process is also not an expedient one when multiple groups such as legal or human resources are involved. This further emphasizes the point an organization's archiving process is a long-term strategy that will change as organizational needs change.

Since the lifespan of messages in an archive can be measured in years, and the need for archiving will increase rather than decrease over time, the solution employed needs to work now, in the year 2020 and beyond. Archiving data in native NSF format is your best option for a viable long-term solution.

---

<sup>8</sup> Referenced from <http://www.financetech.com/featured/showArticle.jhtml?articleID=193004323>



## Appendix A - Definitions

As used in this article, definitions of the following terms:

<i>Term</i>	<i>Definition</i>
<b>Compliance Archiving</b>	Also called, "Journaling." The process of copying a message from the live mail stream or, in some cases, from the primary email data stores (e.g., end user mail files,) to a secure repository for later retrieval.
<b>Email</b>	Electronic mail. Messages transmitted and stored electronically.
<b>Email Archiving</b>	The process of either copying or moving email messages to secondary, but accessible, repositories. There are at least two forms of Email Archiving, "Compliance Archiving" and "End User Archiving."
<b>End User Archiving</b>	The process of moving a message from the primary email data stores (e.g., end user mail files) to user accessible repositories, usually implemented to save space on the primary messaging servers.
<b>NSF</b>	Notes Storage Facility. The term describes the internal structure of Notes databases. Since Notes email messages are stored in Notes databases, this is the native format for Notes email. Notes databases are often, but not always, stored on disk with a <b>.nsf</b> file extension.

