

Understanding Archiving Requirements: Seven Guiding Principles For A Defensible And Compliant Infrastructure Strategy

Mark Diamond

CONTOURAL, INC.

Introduction

From the GC to the CIO down to the storage administrator, there has been no lack of discussion on new rules for managing data and electronic documents. Everything from regulatory compliance such as Sarbanes-Oxley to the Federal Rules of Civil Procedure has made IT aware that they need to be ready to archive more data longer. Yet the most common refrain heard is: "I know I need to do something, just someone tell me what I specifically have to do." How do you cut through the fog, and develop specific technical requirements for saving, managing and deleting data in an archival system? Despite confusion, archiving of data can actually be broken down into fundamental requirements.

Requirement 1: Need To Save All Types Of Data

The most common question asked: What electronic documents and data need to be saved? Only "official records," or documents, files, e-mail, instant messages? The answer to all of these is at various times: yes. Both regulators and especially the courts are expecting organizations to produce (hence have preserved or saved) a wide range of documents. All types of "electronically stored information" (commonly abbreviated ESI) are discoverable. It is important to note that from a regulatory or legal perspective, there is little discrimination of a document based on its medium.

These business and legal drivers create some clear archival requirements IT can follow:

Archiving systems need to be flexible and capable to handle many types of document media – Storage systems used for archiving need to be capable of handling a variety of electronic documents media, including e-mail, instant messages, files, etc

Archiving systems need to be capable of capturing both "official" records as well as "unofficial" documents – Archiving systems need the flexibility to handle input not only from official record and document retention systems, but also electronically stored information from a number of other systems.

Requirement 2: Need To Preserve Large Volumes Of Data

Today many organizations have well-defined and proven processes for retaining paper documents and records. Nevertheless, according to a recent UC Berkeley study, today more than 96 percent of all documents an organization creates or receives are in electronic format. While paper is not going away, it does represent the minority of the medium of documents. Advances in storage technologies through larger-capacity disks have given individual users the ability to save every single e-mail, file, message and other type of document they create, and many do – for many years. This ongoing accumulation both strains storage systems, as

Mark Diamond is President & CEO of Contoural, Inc. Legal information is not legal advice. Contoural provides information regarding business, compliance and litigation trends and issues for educational and planning purposes. Contoural and its consultants do not provide legal advice. Readers should consult with competent legal counsel for professional assurance that our information, and any interpretation of it, is appropriate to each reader's particular situation.

well as represents increased costs and risks in the event of discovery.

Requirements in this area include:

Scalability is Key – Companies commonly underestimate the volume of data likely to go into an archiving system, both from e-mail and other sources. This causes performance bottlenecks across the entire application.

De-duplication is Needed – Many files, messages, etc. have the same content in headers, documents, etc. (called commonality). In order to manage the large volumes of data, an important technical requirement of an archiving system is to be able to recognize this factor and "de-duplicate" the documents.

Archiving Systems Must Be Cost-competitive – Not to be overlooked, archives are likely to contain large amounts of data for a long period of time.

Requirement 3: Need To Search And Retrieve Quickly

If archiving were only about saving data, life would be easier. Archiving is not only about saving, but also of course about getting it back. Organizations looking to deploy archiving need to examine the manner and speed they can retrieve documents, especially for a variety of document media.

The requirements for retrieving documents separate true archiving systems from regular storage:

Searching Content within Documents – Discovery focuses on key words, names or other types of content within a document. Searching for names or other data about a file is insufficient. An effective archival system needs to support complex searches of content within documents.

Search Performance – Extracting documents from the archive, and then searching content is prohibitively slow. What is optimum is to drive these searches from within the archive.

Selectively Retrieve – An effective archiving strategy must encompass the ability to retrieve only targeted files and documents, and not retrieve those that are not relevant.

Case Study: Can't We Just Use Backup Tapes For Archiving?

Many companies have and continue to use backup tapes as the primary mechanism for archiving data for discovery. Backup tapes are designed for copying and restoring large amounts of data, indiscriminately. Archiving has different requirements, namely the ability to preserve, search for and restore specific documents. Backup tapes cannot do this. Furthermore, backup tapes have the additional issue that much of the data contained across multiples tapes may be significantly redundant. Companies that use backup tapes as a primary archiving tool often have to hire expensive outside e-discovery vendors to deduplicate, collate and search through these tapes at considerable expense.

Requirement 4: Need To Preserve And Then Delete

Archiving is often an indeterminate process: when initially archiving data it is not always clear how long the data needs to be saved. More specifically, records preserved for compliance purposes have a defined retention period. That data may then be expired or deleted at the end of that period, unless an organization is compelled to preserve data by a court or a regulator.



Mark Diamond

These "holds" supersede any retention time-lines. These holds often apply to a patchwork of documents, and companies often have an intermixing of data on hold with that which may be legally deleted.

Thus the archiving system must be able to support the following:

Preserve Documents for Prescribed Periods and Then Delete – The archiving system needs to be able to store specific documents for a specified period of time and be able to automatically delete these documents at the end of this period.

Fully Delete Documents – Archiving systems often contain extremely sensitive data, including privacy or confidential information. Hence deleted whole or partial documents should not be recoverable from the medium.

Selectively Preserve Some Documents for Indefinite Periods – The archiving system needs to be able to override the planned destruction of specific or groups of documents in the event of litigation or a regulatory hold.

Selectively Delete Held Documents – While documents may be archived as a group, the archiving system needs to be able to go back and selectively delete documents within that group or across the system.

Saving Records For Lifetime

Save records for a lifetime? In a few cases, yes. A small subset of your records may need to be preserved for anywhere between 30 and 50 years, sometimes longer. For example, OSHA requires that records about workplace accidents be preserved for 30 years after the incident. HIPAA requires that patient data be preserved for the life of the patient plus nine years. Fortunately, these requirements apply to typically a small amount of data.

Requirement 5: Separate Applications From Data Layer

Documents are created from numerous applications, from e-mail systems to databases to Microsoft Word. On many occasions, documents are retrieved long after – in some cases many years – they were initially saved. Furthermore, a few documents such as medical records many need to be saved for literally decades.

Archiving Functionality Separate from Application Layer – Can you access these documents independent of the original application? Archiving systems should be able to understand and manipulate data at the document level. This requires a level of document "intelligence" in the archival system.

Capable of Supporting Open Archiving Standards – The archiving system should be able to support open data formats, including XML, CSV, PDF and others.

Preserve Data for Long Periods – Some or a few documents stored in an archive may need to be preserved for a long period of time. The archiving system should have this capability, or a defined migration strategy.

Requirement 6: Need To Store Data Safely And Reliably

Data in an archive needs to be safe, secure and available. Sensitive data comes in

two types: confidential information includes competitive information, trade secrets, classified information or other types of information which were made available to the wrong party and could be damaging.

One other attribute of sensitive data is immutability – how can you demonstrate that the documents retrieved from the archiving system were indeed what was originally saved in the system, and have not been modified. For example, how can you prove that this e-mail written five years ago has not been changed and was actually written five years ago?

Finally, failure of an archive system could have disastrous consequences. These systems hold critical documents, and any failure could have compliance, legal and business repercussions. Long-term reliability is key.

Additional technical requirements include:

Ability to Encrypt Sensitive Information – While a specific requirement for some regulations, such as HIPAA, encrypting documents is increasingly becoming a best practice for information regardless of regulatory requirement.

Ability to Authenticate Documents – The archiving system should be able to authenticate the creation and integrity of documents in the archive.

Long-term Reliability – An effective archiving system must be reliable. If you don't have long-term reliability, you don't have a long-term archive.

Requirement 7: Need To Control

At the end of the day, archiving is not about saving data. While many discussions about archiving start with how long an organization should save or delete data, while important, this is not the most important factor. Rather, in its most elemental form, archiving is really about controlling your data. Spanning across compliance, discovery, privacy, and cost factors, control is paramount.

Good archiving strategies drive control in the following ways:

- Know what documents, records and data you have;
- Know where your documents are;
- Enable preservation of data when required;
- Ensure deletion of data when you need to;
- Control access and security of your data;
- Do all of these consistently, cost-effectively and compliantly.

It's about control.

These business and legal drivers create some clear archival requirements IT can follow:

Automate Retention and Destruction – The retention and deletion process during the normal course of business (excluding litigation holds) should be a routine, automated process.

Keep an Audit Trail of Retention and Destruction – Archiving systems should keep an audit trail detailing when documents were saved and then deleted.

Hitachi Data Systems recognizes the importance of a robust IT infrastructure for litigation readiness. We are proud to have commissioned Mr. Diamond for this article, which is a condensed version of a NEW two-part whitepaper entitled "Understanding Archiving from an IT Perspective: Seven Guiding Principles for a Defensible and Compliant Infrastructure Strategy."

This new paper is a MUST read for any IT or operational stakeholder; to download your complimentary full-length version, please go to <http://www.hds.com/go/hcap-mcc>.

Hitachi Data Systems develops and delivers Services Oriented Storage Solutions that maximize each customer's return on investment and minimize risk, aligning IT with business objectives and overcoming challenges such as business continuity, disaster recovery, data lifecycle management, and storage consolidation while simplifying IT infrastructures. See www.hds.com.

Please email the author at markdiamond@contoural.com with questions about this article.