Regulatory Compliance and the Hitachi Content Platform

Advancements in technology have changed the way organizations conduct business. Digitization of paper records, pervasive messaging and the explosive growth of unstructured data (file data such as documents, presentations, audio, video) have permeated every aspect of business, bringing with them great benefits as well as increased dependency on these digital assets, which has resulted in new problems. Subsequent calls for increased oversight of specific industries have resulted in new rules aimed at tracking all forms of business communication, including documents, email, email attachments, financial statements, health records and other forms of digital information.

Unlike paper or microfilm, digital information can easily be corrupted, disseminated, copied or altered without authenticity. While the speed of change in data storage technologies continues to accelerate, the length of time required for retention of compliant documents increases, forcing organizations to keep ever growing volumes of data. Storage backup and disaster recovery technologies designed for data availability do not address these issues because the objectives of these technologies are different. System backups are focused on taking a snapshot of all operational information at a given point in time and restoring that snapshot as quickly as possible. The objective of digital archiving is to preserve the “authentic copy” through its retention period and to keep it accessible even as technology advances. Underlying these higher level issues are the infrastructural challenges of maintaining a digital content repository for years to decades and beyond. Storage, server, interconnect and more will change many times over the life of digital assets. Tape formats change, disk technologies evolve and new media types provide compelling cost and performance benefits. Nonstop data growth requires ever
Choosing the right technology infrastructure allows organizations to adapt and evolve over time, helping the modern archive become a familiar extension of the workplace. The Hitachi Content Platform forms the core of a sound compliance solution... simplifying management and reducing costs.

expanding amounts of storage, and regular refreshes of technology become more and more complex and time consuming as data volumes increase. Performing infrastructure overhauls every few years quickly becomes untenable.

Key Requirements for Compliance

It is critical to understand that no individual technology or product implementation will ensure an organization will be judged as compliant by its industry’s governing body or overseeing regulatory agencies. Instead, an organization must implement appropriate business processes, policies and safeguards in conjunction with well-considered technology choices in order to achieve proper compliance. A specific technology or product choice in and of itself will not guarantee compliance, regardless of what certain vendors may claim. This is a very important and often overlooked point. A technology or product choice must be factored into a comprehensive business compliance strategy.

In addition, a common misconception exists that an official certification process is required for compliance with certain high profile regulations such as the U.S. Securities and Exchange Commission’s SEC 17a-3 and 17a-4. Per these particular regulation points, a company deploying anything other than optical disk technology must submit a letter to its designated examining authority indicating it is using a technology (and vendor) that it believes meets the SEC requirements for “electronic storage media.” If the SEC examining authority does not challenge the claim within 90 days, the technology or product choice is deemed acceptable.

That being said, the following list presents some key requirements that are common to most industry regulations and that can be enabled by data storage technology. (Regulations used in compiling the list include but are not limited to SEC 17a-3 and 17a-4, National Association of Securities Dealers or NASD 3010 and 3110, Sarbanes-Oxley Act, Health Insurance Portability and Accountability Act or HIPAA, and Basel II.)

- **Retention.** This includes explicit rules governing the length of time documents must be retained. In some cases, the ability to extend retention periods is required, but the retention period is never shortened.

- **Data Destruction.** Related to retention policy, this dictates when and how specific business documents must be disposed of or destroyed. Storage systems must implement techniques to ensure documents are truly unrecoverable.

- **Authenticity.** During legal investigations or litigation organizations may be called upon to prove that documents have not been altered.

- **Access Controls and Auditing.** This stipulates strict controls over who has access to the original or copies of certain types of documents. Specific data is subject to strict security measures for access and audit trails to provide proof of compliance.
Discovery. During litigation organizations are required to produce business documents within a specified period of time. Most organizations are poorly prepared to produce documents and have become a common target for litigators who make digital business communications, especially email, the target of discovery. Regulations that refer to serialization (SEC 17a-4) and the download of indexes require an IT policy that ensures quick access to documents.

How the Hitachi Content Platform Addresses Requirements

Hitachi Content Platform is a distributed object storage solution offered by Hitachi Data Systems that incorporates highly available enterprise-class storage with intelligent digital archiving software. The list below describes the capabilities of the Hitachi Content Platform with regard to the requirements outlined above.

Retention. “Write once, read many” (WORM) functionality is employed to set a specific file retention period. The retention period can be extended but not shortened. Retention can be set on an object-by-object basis or by selecting related retention policies. Once the retention period has been met, files can be deleted from the system.

Data Destruction. Files can be deleted upon expiration of the retention period, but cannot be deleted when a retention policy is in effect. The Hitachi Content Platform incorporates a “digital shredding” feature that overwrites deleted files with a random pattern, a technique that complies with the United States Department of Defense (DOD) specification 5520.22-M. These actions can be performed on individual objects or a policy can be assigned to automatically govern the deletion of content. All delete actions are logged and the logs can be extracted with the auditing mechanisms in the Hitachi Content Platform.

Authenticity. A digital signature for each incoming file is created utilizing any one of the following hashing algorithms to ensure data integrity: MD5, SHA-1, SHA-256, SHA-384 or SHA-512. The system periodically computes the digital signature and compares it with the original value stored when the file was first archived, ensuring data integrity.

OTHER CONSIDERATIONS

Data Preservation

■ **Provenance** is the first principal of a traditional archival process. Records in archives are arranged according to provenance, a principle that states records of different people or groups should never be mixed.

■ **Original order** dictates that records should be kept arranged in the order in which they were found. Unlike the rule of provenance, the rule of original order can sometimes be broken. For example, records may not have been kept in good order, or may have been kept in no discernible order.

■ **Ingestion** takes place in two stages: appraisal and accession.
  - During appraisal, the organization determines what information sources should be published to the archive. This step is aided by features in the applications responsible for extracting data out of the production systems.
  - Accession is the actual movement of the records into the archive.

■ **Preservation** of records involves both ensuring that the record contains the same information it did when it was archived and ensuring that the record can be viewed using existing technology. It also involves maintaining security and intellectual property rights. The primary responsibility of the preservation layer is to keep the digital record intact. Because digital information is intangible, this is an enormous concern. Each record must be periodically refreshed, and when the hardware that created or stored a record becomes obsolete, the record must be automatically moved to a new device. As data formats become outdated, records must be evolved to support the new standards.

■ **Accessibility** of records: Archives are not data graveyards. A modern archive should be thought of as a component in systems that enforce compliance, improve knowledge transfer and in general enable better corporate governance. By preserving data in a digital archive, the information is readily available and easily managed as needs change.
Solution Profile

Encryption. All data is automatically encrypted using the U.S. National Security Agency (NSA)-approved Advanced Encryption Standard (AES) algorithm before being written to disk. On reads, the data is decrypted and presented back to the requestor in its original format. The encryption and decryption operations are transparent to users and applications.

The encryption key is generated at system installation time and stored internally, eliminating the need for external key management schemes. The encryption key is broken into a number of pieces and distributed among the nodes in the system, ensuring that if a disk or a node were stolen, the data would be unreadable.

Access Controls and Auditing. Administrative access can be restricted to individual IP addresses or a range of allowable addresses. Each access gateway has its own security mechanisms. In addition, all ports not needed for the interfaces are protected by an embedded firewall. The Hitachi Content Platform logs significant events, such as object deletion, so that these actions can be audited.

Discovery. A full content and metadata search and indexing mechanism supporting 370 file formats and 77 languages is available for Hitachi Content Platform. This enables rapid location of documents related to keywords, file properties and custom metadata.

Conclusion

The challenge of building a modern archive goes beyond the need to comply with individual government regulations. The full extent of the challenge is to create a platform that is flexible and can be used in many different environments over a period of time, and yet rigid in requiring that vital information, particularly about the original context of the material, be preserved so that these records can be accessed, authenticated and understood. No single vendor can deliver the full modern archive. The solution involves hardware, software and oftentimes services to develop the unique processes that are needed in a given organization. Choosing the right technology infrastructure allows organizations to adapt and evolve over time, helping the modern archive become a familiar extension of the workplace. The Hitachi Content Platform forms the core of a sound compliance solution as a platform that not only achieves compliance, but also goes far beyond, helping to share information, simplifying management and reducing costs.