

An Evaluation of Hitachi Content Archive Platform

I. Executive Summary

Summary of Evaluation

It is the opinion of Kahn Consulting, Inc. that Hitachi's Content Archive Platform provides a secure and trustworthy platform for the long-term storage of electronic business records. Long-term content integrity, authenticity, and completeness are substantially enhanced by the Content Archive Platform's enterprise-wide storage and intelligent digital archiving software. Assuming that the system is maintained as required, the online nature of the storage system and its capabilities to index and manage infinite volumes of content promote long-term access and retrievability to support business and discovery needs. Furthermore, the Content Archive Platform's ability to protect information from inadvertent or deliberate alteration and destruction can assist organizations to comply with certain information management and compliance requirements.

Evaluation Overview

Kahn Consulting, Inc. ("KCI") was engaged by Hitachi Data Systems to evaluate the company's Content Archive Platform. The purpose of this Evaluation is to assess the product's compliance with general information and records management principles and to gauge its suitability as a platform for the secure, long-term storage of trustworthy electronic business records.

In undertaking this engagement, KCI exclusively relied upon information supplied by Hitachi Data Systems through internal and external documentation, and interviews with Hitachi Data Systems' representatives. KCI did not evaluate the Content Archive Platform in a live or laboratory setting or otherwise field-test any Hitachi Data Systems' products.

Not a legal opinion or legal advice. For all questions regarding compliance with specific laws and regulations seek legal counsel.

WHERE LAW & TECHNOLOGY MEET



October 2006

II. Introduction

Organizations increasingly rely upon electronic records as the “official” records of their business transactions and activities. This reliance mirrors developments in the law, which now allows organizations to use electronic records and signatures for most purposes, in most jurisdictions. However, neither the law nor best practices have altered the need for records to be trustworthy – that is, to be complete, secure, authentic, accessible and to have integrity. The impermanent nature of electronic records can make it difficult to ensure their trustworthiness over the long term.

To address this reality, regulators and other authorities have developed criteria that recognize that electronic business records can easily be changed, deleted, lost or otherwise inadvertently or deliberately manipulated as a result of oversight or fraud. At the same time, the courts can be rigorous when examining the trustworthiness of electronic business records offered as evidence, and may exclude information that is found lacking.

With the passage of a variety of new laws impacting how information is managed, companies are increasingly looking for a storage solution that addresses their business needs but does so in a way that will bolster their value and minimize the likelihood that the information will be successfully challenged as inadequate or untrustworthy.

WHERE LAW & TECHNOLOGY MEET



III. About Content Archive Platform

Hitachi's Content Archive Platform provides an active archive environment for fixed-content with a single online repository that enables protection, search, and retrieval across all content types. The Content Archive Platform is based on a unique SAN-based architecture that leverages Hitachi storage capabilities for availability, performance, and multi-petabyte scalability. The Content Archive Platform provides authentication and secure retention of content to ensure its long-term preservation and accessibility in an active fixed-content archive storage environment.

The Content Archive Platform is designed to enable archiving of fixed content in a manner that:

- Ensures content integrity, authenticity, security, completeness and accessibility over the long term, in accordance with relevant laws and regulations
- Offers fast, online access to content
- Allows integrated searching and indexing of the archive, including search of file contents
- Allows multi-application access through standard communication protocols
- Minimizes the burden of system configuration and management
- Reduces the disruption and expense of media migration
- Supports business continuity and data recovery needs
- Scales horizontally to support multiple applications and content types; and vertically to support continued data growth
- Allows for sub-second search response

Content Archive Platform Architecture. The Content Archive Platform is a complete pre-configured, pre-integrated appliance package that includes hardware and software with configurable storage capacity that is designed to support large, growing archives of fixed-content data. The Content Archive Platform provides a true single cluster with no theoretical limit on the number of nodes it can support, which allows for significant content growth without there being significant negative performance impact. The Content Archive Platform stores files as objects by coupling the file data with the metadata that describes it. The archive solution can maintain all file types, from simple text files to medical image file to multigigabyte database images. The Content Archive Platform itself is not an information or records management application, but a system that works “behind the scenes” of such applications to store the content that they produce. The Content Archive Platform can be connected to and integrated with most software applications through standard communication protocols – SMTP, NFS, CIFS/SMB, HTTP, HTTPS, and WebDAV.

To meet its design goals, the Content Archive Platform incorporates several unique features. This Evaluation focuses on those features designed to meet general criteria for the secure, trustworthy long-term storage of records and business information.

Content Archive Platform Features

Single Level of Management

Regardless of the number of nodes, amount of capacity or size of content the Content Archive Platform provides the capability to manage the environment as a ‘single system’ through its SAN + Redundant Array of Independent Nodes (SAIN) architecture and scalable file system. As the need to store digital content continues to grow at a rapid pace, the ability to manage the environment as

WHERE LAW & TECHNOLOGY MEET

KAHN
CONSULTING INC.

a single system becomes very critical to support easy and fast accessibility to content.

Full Content Searching

The Content Archive Platform has the capability to perform expeditious search and comprehensive searches by name, file attributes, metadata, and file content across applications and document types. For an archive to be searchable, the Content Archive Platform cluster must have one or more search nodes. Search nodes manage a distributed index of all the archived files in the cluster. For fast retrieval of query results, the search engine maintains an index, which is based on file data and metadata. The ability to search file content is becoming more critical to corporation to be able to respond quickly to discovery and audit requests. File names or metadata may not always contain the key words that are requested during discovery, thus the ability to search the file content becomes very instrumental in locating all files that may be related to the matter. The Content Archive Platform has the capability to support 77 languages and over 270 file types.

Open Standard Interface

The support of SMTP, NFS, CIFS/SMB, HTTP, HTTPS and WebDAV communication protocols allows for most applications to work with the Content Archive Platform. The Content Archive Platform also provides an API for search that enables tighter integration with specific applications. Implementing the Content Archive Platform solution is faster and less costly with the open standard architecture than with a single API solution. Use of standard protocols ensures that any part of the archive solution can be replaced, mitigating technology obsolescence in the future.

Self Healing

The Content Archive Platform has the capability to monitor and detect the status of each node within the archive. The protection policy within the archive maintains the integrity of the objects by initiating repairs automatically immediately after any failure. All files stored in a cluster are subject to a tolerable-points-of-failure data protection policy. If a node or logical unit fails, the Content Archive Platform automatically redirects operations to other nodes. It also signals the protection policy to begin creating additional copies of the data and metadata on a separate storage device. The protection policy ensures that archived data is always available and protected.

SAN + RAID (SAIN)

The Content Archive Platform is built on a SAN + RAID in an array of independent node architecture model. It is delivered as a pre-configured, integrated package that consists of servers, software, and Hitachi's TagmaStore Workgroup Modular Storage® systems. It can be configured in 5 and 10 terabyte storage units. The base configuration consists of a 4 server-node cluster that is scalable in 2.5 TB and 5TB increments.

Easy Integration with Operations

The Content Archive Platform can be monitored and managed from any Simple Network Management Protocol (SNMP) compliant network management station. Critical event information such as disabled node, disabled drive, policy violations, time synchronization problems and the like are made known through the Web Admin User Interface. The Content Archive Platform also has the capability to respond to SNMP GET, GET NEXT, and SET directives. Operations personnel can be notified when critical events occur via email, SMS, pager, and other methods through the network management system. This capability eliminates cost associated with managing the day-to-day monitoring activities for the Content Archive Platform.

WHERE LAW & TECHNOLOGY MEET

KAHN
CONSULTING INC.

Content Ingestion

Large quantities of content can be rapidly ingested from an application through the use of optimized standard communication protocols – SMTP, NFS, CIFS/SMB, HTTP, HTTPS and WebDAV, as opposed to ‘one API fits all’ approach. The capability to select the access gateway depending on the producing application, allows for optimization during the transfer process of data to the archive storage location.

Remotely Serviceable

The Content Archive Platform does not require immediate onsite access for all hardware failure. The Content Archive Platform has remote ability to run diagnostics, install patches and upgrades through IP modem or VPN.

Linear Scalability

As data continues to grow, additional nodes and external storage can be added to the Content Archive Platform. The archive capacity can grow easily from hundreds of gigabytes to terabytes and petabytes. By distributing its processing across the network of nodes in the archive, the system can scale as the archive or numbers of users grows in size.

Additional Features

The Content Archive Platform offers a variety of additional features that support its use as a platform for the secure, trustworthy long-term storage of electronic business records, as will be explored in detail in the next section.

WHERE LAW & TECHNOLOGY MEET



IV. Content Archive Platform Capabilities

Electronic business records and other fixed content must be stored and managed in a trustworthy manner. There is little point in expending resources to store content if it cannot be relied upon for operational, legal, compliance and historical purposes. Trustworthiness is most accurately thought of as a quality that results from the sum total of the people, procedures, environments, strategies, and technologies used throughout the lifecycle of a business record. Trustworthiness suggests that a court, regulator, auditor – and the organization itself – can trust and rely upon the content of a record.

The system used to store and manage digital information plays a large role in ensuring the trustworthiness of the stored information. This part of the Evaluation is divided into sections that describe capabilities that are desired in storage systems; explain why each capability is desired; and assess the Content Archive Platform's compliance with each capability.

1. Long-term Content Integrity (Preventing Alteration)

Desired Capability. Electronic records and business information should be protected from inadvertent or deliberate alteration. A system that protects records from alteration can minimize the likelihood that the authenticity of electronic records will be successfully challenged in court or by a regulator.

Information Management Principle. Information is said to have integrity if it can be demonstrated that it has not been altered since it was created or archived. Unlike paper-based information, which has inherent features that deter alteration (such as the physical bond between ink and paper); the alteration of most digital information in its native form is easily accomplished without detection. Business best practices and many laws and regulations require digital information to have integrity.

Content Archive Platform Capabilities. Objects written to the Content Archive Platform are stored in a 'write once, read many' ("WORM") storage environment. Once a record is stored on WORM it cannot be deleted or altered prior to the end of its retention period assigned by the controlling application. WORM records cannot be altered anytime throughout their entire lifecycle up until their final disposition. Protection and retention policies and various kinds of metadata are also used to ensure the integrity of archived objects.

The Content Archive Platform policies prohibit renaming a file, modifying content of a file, overwriting a file, deleting a file prior to expiration of the retention period, deleting a directory that contains files, modifying metadata other than ownership, permissions, retention periods (lengthen only), shred settings, shortening retention period, and deleting metadata files or directories. The policies embedded in the Content Archive Platform were designed to support long-term content integrity for archived objects.

- 1) **Hashing¹.** All information sent to the Content Archive Platform is processed by a user selectable hashing algorithm. The Content Archive Platform supports SHA-1, 256, 384, 512, MD5 and RIPEMD-160 hashing algorithms. The algorithm produces at the binary level a fixed-length "fingerprint" of the object. The fingerprint is used to determine when the object has changed and generates a new fingerprint for the object.
- 2) **Validation.** The data and metadata for each object are stored in multiple locations across the nodes to ensure ongoing data integrity and validation. The storage manager functionality in the Content Archive Platform performs integrity checks on the data continuously.

WHERE LAW & TECHNOLOGY MEET

KAHN
CONSULTING INC.

- 3) **Deliberate Alteration.** Once an object is stored in the Content Archive Platform, it cannot be altered even through the controlling application. When alteration occurs in the controlling application, the Content Archive Platform will recognize that object as a new object and will generate a new fingerprint. This process prevents the deliberate alteration of objects that have been previously stored.
- 4) **Inadvertent Alteration.** The authentication policy performs two functions – detecting authentication violations and metadata discrepancies. If violations or discrepancies occur, the authentication policy also has built in repair capabilities. The authentication policy regenerates the digital signatures for each archived file from the file data only. An additional signature is generated from the entire file, including the header. After regenerating the signatures, the policy checks that the regenerated signatures match the corresponding signatures in the primary metadata. The policy also checks the UID, GID, file permission, and retention settings to ensure they have not been inadvertently altered.

2. Recording and Storage Process Integrity and Authenticity (System Trustworthiness)

Desired Capability. When organizations archive electronic records and business information for future use, the reliability and integrity of the initial recording and storage process should be validated.

Information Management Principle. Information cannot be relied upon unless there is assurance that the information was recorded in a manner that reflects the form and content of the information as it was originally created.

Content Archive Platform Capabilities. To ensure that the object maintains its integrity when transferring from the archiving application to the Content Archive Platform, while ingestion occurs, the archiving application computes the cryptographic hash in parallel to ingestion. Once the archiving application completes the input, the hash has been totally computed. Once ingestion is complete, the Content Archive Platform will have also computed a hash. The archiving application then compares the Content Archive Platform hash to what it has computed.

3. Content Accessibility

Desired Capability. Organizations should be able to access information in a timely, trustworthy, and cost-effective fashion at any time during the information lifecycle.

Information Management Principle. Information that cannot be readily found and accessed is of marginal utility. In the short term, responding to regulator and court requests for information or documents must often be completed in a short timeframe. In the long term, the limited lifespan of every digital storage medium; data corruption; heterogeneous storage formats; technological obsolescence; and other factors threaten the long-term accessibility of electronic information and records.

Content Archive Platform Capabilities. The Content Archive Platform uses magnetic disk to provide faster access times than other media that is typically used for the archiving of electronic records. Additionally, the Content Archive Platform architecture enables archived information to remain online and accessible without performance degradation. Conversely, systems that rely on removable media, such as optical disk, typically employ a staged system where only a certain number of disks remain in the storage device for immediate access. In this regard, the Content Archive Platform may provide faster and more cost effective access to information than other kinds of storage systems designed for archiving content. The use of magnetic disk also reduces the need

WHERE LAW & TECHNOLOGY MEET

KAHN
CONSULTING INC.

for human intervention (e.g. to load media), which eliminates time delays and labor costs associated with accessing archived information. The Content Archive Platform's capabilities in this area may also allow for faster access to information in support ongoing business operations or to respond to requests for information by regulators or in the context of litigation.

The Content Archive Platform allows access to archived content through NFS, CIFS/SMB, HTTP, HTTPS, SMTP, and WebDav gateways. These gateways can be used to access the archive with a web browser, third-party applications, Windows Explorer, or native DOS or UNIX tools. This functionality may be useful in the context of discovery to ensure a comprehensive and complete search of all relevant content without identifying the controlling application to perform the search.

4. Business Continuity and Disaster Recover

Desired Capability. Standard disaster recovery techniques require that data is stored in at least two physically separate locations.

Information Management Principle. Data that does not exist in two or more places can be permanently lost if the device or facility housing the data is damaged or destroyed. Thus, there is a need for organizations to copy important data to different physical locations for backup, disaster recovery, and business continuity purposes.

Content Archive Platform Capabilities. The Content Archive Platform can be configured to continuously replicate the contents to a physically separate cluster. System administrators define the number (between 2 and 4) of replica copies that are maintained depending on the value of the content. This capability assists the organization in meeting its long-term preservation needs for ensuring archived data is available in the event an object is no longer available for whatever reason.

The Content Archive Platform also provides a backup and restore Network Data Management Protocol (NDMP) standards based interface to allow offline, off-premises backups and restore to support disaster recovery and business continuity in the event of a physical disaster. NDMP access protocol provides open, standards-based backup and recovery to removable media of choice. Full, incremental, and differential backup sets are performed.

The Content Archive Platform utilizes Archive Object Package (AOP) format that includes compression, digital signature, and encryption when replicating and backing up content. The AOP for data object consists of three files. One file containing the object data, one file containing the object metadata, and one file that identifies the other two files.

5. Resistance to Deletion and Overwriting

Capability. Storage systems designed to store highly sensitive and regulated information should offer the capability to protect information from being inadvertently or deliberately deleted or overwritten.

Information Management Principle. In order to satisfy certain business requirements, laws, regulations and other criteria, electronic records may need to be stored in a fashion that ensures that they cannot be deleted or overwritten.

Content Archive Platform Capabilities. Each data file and directory in the archive stipulates the period of time that an object in the archive must be retained through the retention policy functionality. The retention metadata is associated with the object and maintained through its lifecycle. Once the retention has been established for the object, it cannot be deleted or overwritten

WHERE LAW & TECHNOLOGY MEET

KAHN
CONSULTING INC.

prior to the expiration of the retention period. If an application or user attempts to delete a file prior to the expiration of the retention period, the retention policy, running continuously, prevents it from doing so. The Content Archive Platform also will not allow a ‘work around’ deletion to occur by shortening the retention period. The retention policy only allows for extending a retention period on an archived object by users that have been granted secure login privileges based on corporate policies, thus preventing the user or application from changing the retention period to a shorter time in order to execute a delete. This functionality prevents unauthorized deletion and overwriting of archived information, which is an important component of managing long-term preservation of information.

6. Records Retention

Capability. A storage system designed for the long-term storage of electronic records should offer records retention functionality.

Information Management Principle. Laws, regulations, standards, and practices require organizations to retain specific types of information for specific periods of time. Organizations retaining records in electronic form require storage systems that enable them to designate retention periods for electronic records and dispose of records at the end of their lifecycle.

Content Archive Platform Capabilities. Record Retention is set by the controlling applications or manually by a records administrator at a directory or file level. The assigned retention period is stored with the object and the retention policy ensures deletion or overwriting does not occur before the end of its designated retention period. The Content Archive Platform allows for two methods to ‘start the clock’ on retention. One method is assigning a finite fixed time period that will start the retention clock on the day of ingest (after being stored in the system). Event based retention is the second method allowed by the Content Archive Platform. The content is originally set as infinite and when the event occurs, the retention is changed to finite and the clock starts. Retention can be set as short as 0 minutes after ingest to infinite based on the policies set by the controlling application or directory.

Retention periods cannot be shortened once an object is stored in the Content Archive Platform. However, if there is a need that necessitates lengthening the original retention the Content Archive Platform provides the capability to modify the retention period.

The Content Archive Platform supports the use of a “HOLD” setting to prevent files from being deleted until they are “UNHELD” for those files that do not have a permanent retention period. This functionality is used most often during discovery, litigation or audits to ensure pertinent files to the matter are not inadvertently deleted when normal retention criteria has been satisfied.

7. Information Security

Desired Capability. Storage systems should provide information security controls and capabilities that protect the system and its contents from alteration, corruption, inaccessibility, loss, compromise of confidentiality and privacy, and other damaging events.

Information Management Principle. Organizations manage and store valuable information that must be protected. In some cases, confidentiality must be maintained, and in other cases privacy protection is a legal requirement. Security is a complex process that involves many different procedures and technologies, but it is fundamental to an organization meeting its information management goals and obligations.

Content Archive Platform Capabilities. The Content Archive Platform offers a variety of controls and techniques designed to secure the system and its contents, as follows:

WHERE LAW & TECHNOLOGY MEET

KAHN
CONSULTING INC.

- 1) **Architecture.** The Content Archive Platform is controlled through the administrative console by granting access to specific Internet Protocol (IP) addresses. This architecture makes it difficult for an attacker without access information contained in the archive. ²
- 2) **Access controls.** Access is controlled through verification of IP addresses and passwords that are established by the administrator. Read and write access levels can be restricted by the administrator. Secure login is used to grant privileges to appropriate users that require access to system controls to manage the archive.
- 3) **Administration.** The Content Archive Platform can be configured to disallow remote administrative access. When remote administrative access is disallowed, administrative access requires direct physical connection to the Content Archive Platform cabinet that is located onsite in a physically secure location.
- 4) **Application Access.** A controlling application's access to the Content Archive Platform can be restricted to specific IP addresses. Therefore, access can only be granted to an application by the administrator.

8. Content Completeness and Authenticity

Desired Capability. Information storage systems should retain information in manner that preserves its complete content, physical form, layout, and metadata – especially that metadata indicating origin and provenance.

Information Management Principle. An electronic record is said to be authentic if it is in fact “what it purports to be.” That is, the source or origin of the e-record can be reliably demonstrated. This often requires proof of who generated an e-record, and who controlled it at a certain time. In addition, an electronic record should remain in a complete and accurate form throughout its lifecycle in order to be considered trustworthy, and to satisfy a variety of business and legal requirements.

Content Archive Platform Capabilities. Information sent to the Content Archive Platform by a controlling application is retained by the Content Archive Platform in a manner that preserves all of the object's original qualities, such as the file format, physical appearance, content, metadata and policy information. The authentication policy regenerates the digital signatures to ensure that the content of the archived file matches its original digital signatures throughout its lifecycle in the archive. Violations can be detected and repaired from an existing good copy of the file. In this manner, the Content Archive Platform provides features that enable electronic records to be stored in an original, complete, and authentic manner.

9. Records Disposition

Desired Capability. Records Management solutions should provide the capability to properly dispose of information once it is no longer needed.

Information Management Principle. Disposition is the final lifecycle stage of most information. Although there are relatively rare cases where information must be retained in perpetuity for historical and other purposes, over time the vast majority of information ceases to be of value to an organization. In the digital world, it can be difficult and expensive to ensure that electronic information is properly disposed of. This can lead to situations where files are not properly disposed of and unwanted files are recovered or recreated in the course of litigation.

Content Archive Platform Capabilities. The Content Archive Platform ensures that the retention

WHERE LAW & TECHNOLOGY MEET

KAHN
CONSULTING INC.

period of the object has expired before deletion can occur. If the retention period has not been satisfied, then deletion can't occur under any circumstance. The Content Archive Platform supports receiving a deletion request from a controlling application or by pre-setting a value to automatically delete an object when the retention period has expired.

After deletion the objects are hidden from the archiving application and placed into a queue to be shredded. The Content Archive Platform supports the use of digital "shredding" techniques that conform to the US Department of Defense 5220.22-M standard for disk sanitization. This method overwrites the physical section of a disk drive containing the object to be deleted seven times with pseudo-random information, thereby making recovery of that object nearly impossible, even using advanced recovery techniques.



WHERE LAW & TECHNOLOGY MEET



CONSULTING INC.

IV. About Kahn Consulting

Kahn Consulting, Inc. (KCI) is a consulting firm specializing in the legal, compliance, and policy issues of information technology and information lifecycle management. Through a range of services including information and records management program development; electronic records and email policy development; Information Management Compliance audits; product assessments; legal and compliance research; and education and training, KCI helps its clients address today's critical issues in an ever-changing regulatory and technological environment. Based in Chicago, KCI provides its services to Fortune 500 companies and government agencies in North America and around the world. Kahn has advised a wide range of clients, including Time Warner Cable, Ameritech/SBC Communications, the Federal Reserve Banks, International Paper, Dole Foods, Sun Life Financial, Kodak, McDonalds Corp., Hewlett-Packard, United Health Group, Prudential Financial, Motorola, Altria Group, Starbucks, Mutual of Omaha, Merck and Co., Cerner Corporation, Sony Corporation, and the Environmental Protection Agency. More information about KCI, its services and its clients can be found online at: www.KahnConsultingInc.com.

V. Endnotes

¹ Industry standard hashing algorithms such as those used by HCAP operate in such a manner that the likelihood of two pieces of different information resulting in the same hash value is extremely low statistically.

² In most cases, a skilled attacker or trusted insider can circumvent even the strictest security controls and mechanisms within any information system, given enough knowledge, resources, and time.

Entire contents © 2006 Kahn Consulting, Inc. ("KCI"). Reproduction of this publication in any form without prior written permission is forbidden. KCI or Hitachi shall have no liability for errors, omissions or inadequacies in the information contained herein or for interpretations thereof. The reader assumes sole responsibility for the selection of these materials to achieve its intended results. The opinions expressed herein are subject to change without notice. All rights reserved. www.KahnConsultingInc.com info@KahnConsultingInc.com 847-266-0722

WHERE LAW & TECHNOLOGY MEET

KAHN
CONSULTING INC.