

# An Evaluation of EMC Centera Governance Edition

Managing Electronic Records for Compliance and Corporate Governance

## I. Executive Summary

### Summary of Evaluation

It is the opinion of Kahn Consulting, Inc. that EMC's Centera™ Governance Edition ("Centera") provides a compelling platform for the trustworthy storage of electronic records and other digital information required for compliance and governance purposes. By protecting the integrity, reliability, accessibility, and accuracy of information, Centera can play an important role in helping organizations retain and manage information in manner that will better ensure its admissibility, promote its evidentiary strength, and support governance processes. By providing long-term content accessibility; by supporting records retention, preservation, and destruction functionality; by facilitating search and retrieval of information; by protecting content from alteration and unauthorized deletion; by verifying the accuracy of information during the recording process; and by supporting disaster recovery and information security needs, Centera promotes the authenticity and trustworthiness of electronic records and digital evidence. Furthermore, Centera's capabilities can assist organizations to comply with the requirements of laws and regulations governing information management, such as the Sarbanes-Oxley Act<sup>1</sup> or the Basel II Accord<sup>2</sup>.

### Evaluation Overview

Kahn Consulting, Inc. ("KCI") was engaged by EMC Corporation ("EMC") to evaluate the company's Centera Governance Edition storage platform ("Centera").<sup>3</sup> The primary purpose of this Evaluation is to assess the product's utility as a platform for the retention, management, securing, and retrieval of electronic records and other digital information as required for compliance and corporate governance purposes. In conducting this Evaluation, KCI has assessed Centera functionality against criteria derived from broad legal and regulatory requirements and best practices for the management of electronic records. Retaining and managing digital information in manner that will satisfy the courts and regulators depends on a proper program of technology, people, and technical and procedural controls. This Evaluation assesses the value that Centera may bring to such a program.<sup>4</sup>

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

WHERE LAW & TECHNOLOGY MEET

**KAHN**  
CONSULTING INC.

January 2006

## II. About Centera

### Overview

EMC's Centera Governance Edition product ("Centera") is designed to provide a long-term storage solution for the retention, management, securing, and retrieval of electronic records. Electronic records include financial spreadsheets, word processing documents, e-mail messages, digital images, and many other types of information that must be kept for business, operational, legal, compliance, and/or historical purposes.

EMC designed Centera to provide evidentiary benefits while leveraging the economic and functional benefits of magnetic disk as part of an enterprise Information Lifecycle Management ("ILM") strategy.

Centera is designed to enable the storage of electronic records in a manner that:

- Ensures record integrity, authenticity, security, completeness and accessibility over the long term, in accordance with relevant laws and regulations
- Supports the production of records and information during electronic discovery
- Offers fast, online access to electronic records
- Minimizes the burden of system configuration and management
- Reduces the disruption and expense of media migration
- Supports business continuity and data recovery needs
- Allows storage repositories to grow in a non-disruptive, flexible manner
- Integrates with existing information and records management applications

### Architecture

Centera is an integrated combination of software and off-the-shelf hardware components sold as an appliance with an expandable storage capacity. Centera itself is not an information or records management application, but rather an online information repository that works transparently "behind the scenes" to retain, protect, and retrieve the content produced by such applications. Centera can be connected to and integrated with a broad range of software applications ("controlling applications") within multiple markets, including, but not limited to: medical imaging; e-mail archiving; enterprise content management; records management; e-learning; audio and video management; workflow; and so on. To expedite deployment and integration, EMC has an open application programming interface (API) available for partners to integrate Centera with many applications in a variety of industries. In addition, the "Centera Universal Access" appliance provided by EMC enables customized applications (i.e., those that are not commercially available by a software vendor) to use Centera.

To meet its design goals, Centera incorporates several unique features. This Evaluation focuses on those features designed to meet general criteria for the secure, long-term storage of trustworthy electronic records for compliance and governance purposes.

WHERE LAW & TECHNOLOGY MEET

**KAHN**  
CONSULTING INC.

## Features

### Content Addressing

EMC's Centera uses a data access paradigm known as "Content Addressing." Content Addressing is a method for storing, accessing, and authenticating a digital file, document, or image (collectively referred to as "digital object" throughout this Evaluation). Content Addressing creates and uses secure alphanumeric object descriptors derived from the content of the object itself. This "content addressing" method differs from traditional "location addressing" methods where digital objects are stored and accessed based on their physical or logical location within the storage system. With Content Addressing, the storage and retrieval of content occurs independently of its physical storage location, and no URLs, file structures, or pathnames are used at all. Content Addressing also offers several evidentiary and compliance benefits, as described below.

### Redundant Array of Independent Nodes (RAIN)

Each Centera includes anywhere from 4 to 128 "nodes" that provide 2.2 to over 180 usable terabytes of data storage. Each node is an independent unit comprised of a motherboard, a processor running Centera software, and four magnetic disk drives. The ability of each interconnected node to process and store data independently reduces reliance on a central system (thus providing greater reliability), and supports advanced data protection, replication, and recovery, as described in detail below.

### Content Protection Mirroring (CPM) and Content Protection Parity (CPP)

Centera provides two user configurable methods for continuously protecting against the loss of digital objects in the system due to data corruption, device failure and so on. The first method, Content Protection Parity (CPP), automatically splits files larger than a certain size<sup>5</sup> into six data fragments and a seventh "parity" fragment. Each of the seven fragments is stored on separate hardware nodes. If any one node were to fail, the missing fragment stored on the failed node can be automatically recreated using any of the remaining five fragments and the parity fragment.<sup>6</sup>

The second method, Content Protection Mirroring (CPM), automatically creates two physical copies of an object on two separate nodes within the system. Using this method, there are always two complete copies of each object within the system; copies that can subsequently be used for data recovery and record regeneration purposes. Unlike traditional mirroring methods, CPM mirroring is done at the object level, rather than at the volume or partition level.

Data protection schemes such as CPP and CPM help to support the long-term storage of electronic records as described below in greater detail.

### Data Regeneration: "Self Healing"

Centera continuously monitors each drive and node for faults in either the node, the drive, or in the stored objects themselves. If a fault is detected in a disk or node, that disk or node is isolated, and its objects automatically are regenerated from the object's mirrored twin or parity fragments to healthy drives or nodes. Similarly, if corruption is found within an individual object, that object is "regenerated" using the mirrored copy or the parity fragments (depending upon the method originally used to store the file). These operations are completed automatically and without disruption to the overall functioning of the Centera system. However, the system is configured to

WHERE LAW & TECHNOLOGY MEET

**KAHN**  
CONSULTING INC.

automatically notify a system administrator when a disk or node failure has occurred. The repair and rebuilding of stored data is an important capability of a long-term storage system.

### Remote Replication and Disaster Recovery

Centera can be configured to asynchronously replicate, over a standard IP network, objects stored on a Centera system (or systems) to a Centera system (or systems) located in different geographical locations. In the event of a disaster or other event resulting in loss of or damage to stored objects, replicated objects can be restored from one Centera system (or systems) to another Centera system (or systems) in a different geographical location. Centera provides four options (or “topologies”) for asynchronous remote replication and restoration.

Using the first option (referred to as “unidirectional”), when an object is stored on the primary Centera system, it is automatically copied to the replicated Centera system at the remote location. In the event of a disaster, objects can be accessed from the remote location.

Using the second option (referred to as “bidirectional”) objects written to any Centera system that is part of the network will be replicated to the other Centera system/s in the network.

The third option (referred to as “star”) allows multiple Centera appliances to replicate data to a primary Centera. This might be used, for example, in a situation where multiple field offices have Centera installations that they wish to backup to a central data center.

The fourth option (referred to as “chain”) allows objects to be replicated on both a secondary and a tertiary Centera appliance.

The remote replication capability offered by Centera helps to ensure that organizations will be able to access their information assets in the event of a disaster - a critical capability for all organizations today.

### Virtual Pools

Centera enables organizations to store information from multiple applications on a single Centera system through a feature EMC describes as “Virtual Pools.” A virtual pool is a set of information that is written by a specific controlling application. After virtual pools are established, controlling applications can perform operations such as read, write, and delete on specific pools of information, rather than to the entire Centera system. This might be used to selectively replicate only certain pools of information, or to search only the data created by a particular application, for example. In addition, pools can be restored on a pool-by-pool basis using Centera’s remote replication capabilities as described above.

## III. Centera Capabilities

Electronic business records must be stored and managed in a trustworthy manner. Trustworthiness is most accurately thought of as a quality that results from the sum total of the people, policies, procedures, environments, strategies, and technologies used throughout the lifecycle of a business record. The technology used to store and manage digital information plays an integral role in ensuring the trustworthiness of the stored information. As stated in the Federal Rules of Evidence, evidence can be authenticated by “evidence describing a process or system used to produce a result and showing that the process or system produces an accurate result.”<sup>7</sup>

WHERE LAW & TECHNOLOGY MEET

**KAHN**  
CONSULTING INC.

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 Centera's compliance with each capability.

## Accessibility

**Desired Capability.** Organizations should be able to access information in a timely and cost-effective fashion at any time during the information lifecycle. As stated by one court, “[u]tilizing a system of record keeping which conceals rather than discloses or makes it unduly difficult to locate” records may be considered the equivalent of destroying records.<sup>8</sup>

**Information Management Principle.** Information that cannot be readily found or accessed is of little use to an organization. It may also be a source of legal risk, as responding to a regulator or a court request for records must often be completed in a short timeframe. In one case, a firm was fined \$10 million for failing to produce information in a timely fashion.<sup>9</sup>

**Centera Capabilities.** Unlike solutions based on optical disk, tape, and other storage formats typically used for the archiving of electronic records, Centera uses magnetic disk. Magnetic disk can provide faster access times than certain other media. In addition, the Centera architecture is designed to enable an entire archive of information to remain “online,” searchable, and accessible without significant 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, Centera may provide faster and more cost effective short-term access to information than other kinds of electronic records storage systems. In addition, unlike tape libraries and other approaches, access to information stored on Centera does not require human intervention (i.e. to load media, and so on), which works to minimize time delays and labor costs associated with accessing archived information. Centera's capabilities in this area not only support business operations, but may also reduce costs associated with finding and producing information in the context of electronic discovery.

## Retention Period Coding

**Desired 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 specified periods of time. Organizations retaining records in electronic form require software applications and storage systems that enable them to designate retention periods for electronic records and destroy records at the end of their lifecycle.

**Centera Capabilities.** Centera enables controlling applications to designate record retention periods. This information is stored with object throughout its lifecycle and protects the object from being deleted or overwritten before the end of its retention period. After an object is stored in Centera, its retention period cannot be shortened. The retention period can, however, be lengthened if changing retention criteria, electronic discovery requirements, or other factors require an extension of the original retention period.

Centera offers additional retention capabilities, including:

WHERE LAW & TECHNOLOGY MEET

**KAHN**  
CONSULTING INC.

157 LEONARD WOOD NORTH • HIGHLAND PARK IL • 60035  
PHONE: 847.266.0722 • FAX: 847.266.0734 • EMAIL: INFO@KAHNCONSULTINGINC.COM

- 1) **“Retention Classes.”** Centera can be configured so that retention periods are assigned to groups of records. This capability allows administrators to apply retention periods to an entire class of records, rather than applying retention periods on a record-by-record basis. Administrators using this capability will need to ensure that records requiring different treatment are properly indexed, coded, and retained.
- 2) **Mandatory Retention.** Centera can be configured in such a way that information cannot be written to Centera unless a retention period has been assigned.
- 3) **Default Retention Periods.** Centera can be configured to automatically assign retention periods to information for which a retention period has not otherwise been assigned by the controlling application.
- 4) **Event-Based Retention.** Centera can be configured to assign retention periods to information based on business events that are communicated to Centera by the controlling application. This capability supports the need that many organizations have to determine record retention periods based on specific events, such as the end of the contract, closing of a customer account, or payout of a life insurance policy, for example.

## Search and Retrieval

**Desired Capability.** Electronic records storage systems should provide efficient and effective search and retrieval capabilities.

**Information Management Principle.** Today, more than ever before, organizations must be able to quickly and cost-effectively find and retrieve specific electronic records and information. This requirement may derive from the legal discovery obligation associated with lawsuits, investigations, and audits, or it may simply derive from a need to operate the business more effectively.

**Centera Capabilities.** Through the use of Centera’s “Virtual Pools,” the search and retrieval of information can be limited to only records created by a specific controlling application in a certain timeframe. In the context of e-discovery this can help to limit the time and expense associated with searching for responsive records and information.

In addition, the Centera “Seek” function enables sophisticated searching and retrieval of objects through the use of an index that contains many different types of metadata about each object - including customized metadata.<sup>10</sup> Centera creates its own metadata (such as the date and time the information was stored, its retention period, etc.), and metadata can also be created as required by the controlling application to serve future search and retrieval needs. This metadata is stored in XML format – a non-proprietary data format that is both human and machine readable. In addition, the XML file can be used by Centera independently from the controlling application that generated the metadata - a feature that may be useful in instances where the controlling application is no longer available. These capabilities are designed to support the search and retrieval of information for compliance, governance, discovery, and business purposes.

## Preservation of Information

**Desired Capability.** Electronic records storage systems should support the preservation of information in the face of lawsuits, audits, and investigations.

WHERE LAW & TECHNOLOGY MEET

**KAHN**  
CONSULTING INC.

**Information Management Principle.** In instance where organizations face lawsuits, audits, investigations, and other formal proceedings, they may be required to preserve any information relevant to the proceeding and protect it from disposition or alteration.

**Centera Capabilities.** Centera is capable of assigning a “litigation hold” to an object, which prevents that object from being deleted even if its previously-applied retention period has expired. This capability could be used by administrators to protect information subject to preservation requirements from being deleted and can help an organization fulfill its discovery obligations.

## Preventing Alteration

**Desired Capability.** Electronic records 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 has integrity if it can be demonstrated that it has not been altered and remains accurate since it was created or archived. Business best practices and many laws and regulations require digital information to have integrity.

**Centera Capabilities.** The Centera Content Addressing System works to prevent the inadvertent or deliberate alteration of information, as follows:

- 1) **Hashing.** All information sent to Centera by a controlling application is processed by a “hashing” algorithm. This algorithm processes a digital object at the binary level to produce a fixed-length digital “fingerprint” of the object. This fingerprint is the unique by-product of that digital object, and that digital object always creates the same fingerprint when it is processed by the algorithm. However, if an object is changed in any way, a new fingerprint is calculated by the hashing algorithm and stored in Centera. This process in effect ensures that any alteration to an object is detected.
- 2) **Content Addressing.** The object’s fingerprint is used by Centera (along with additional information) as that object’s Content Address (CA). The CA is stored and used by Centera to access and authenticate the object throughout its entire lifecycle. The CA is unique to that object, does not change over the life of the object, and is not dependent upon the location of the object in the Centera system.<sup>11</sup>
- 3) **Validation.** Each digital object’s CA is recalculated during all significant interactions between Centera and the controlling application, and is also continuously calculated by built-in Centera data validation utilities that run perpetually within each node, comparing each object’s current CA to its original CA.
- 4) **Preventing Deliberate Alteration.** There is no direct access to the files in Centera, as described in detail below. Files can only be altered within a controlling application after retrieving the object from Centera. If a user retrieves an object from Centera and alters that object, when the altered object is sent back to Centera, Centera’s hashing algorithms automatically will calculate an entirely new Content Address and will manage the new, revised object separately from the original object. The original object will continue to be stored in its original, unaltered state. In this manner, Centera works to prevent the deliberate alteration of objects stored within it.<sup>12</sup>

WHERE LAW & TECHNOLOGY MEET

**KAHN**  
CONSULTING INC.

- 5) **Preventing Inadvertent Alteration.** If an object is corrupted or otherwise altered inadvertently, this change will be revealed by the automatic, ongoing comparison of Content Addresses, and the altered file will be automatically replaced with the mirrored file, or rebuilt from the parity fragments. EMC calls this process “organic regeneration.” The process of comparing Content Addresses also occurs when a Centera file is requested (or “read”) by a controlling application. In this way, Centera protects against inadvertent alteration of stored data.

## Preventing Deletion or Overwriting

**Desired Capability.** Storage systems designed to store electronic records should offer the capability to protect those records from being inadvertently or deliberately deleted or overwritten. During the audit, investigation, or litigation, improper alteration or deletion (i.e., “spoliation”) of information can result in criminal charges, severe fines, penalties, and other negative consequences.

**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. From an evidentiary perspective, such a capability helps to demonstrate record integrity and preempt attacks on record trustworthiness.

**Centera Capabilities.** A controlling application can stipulate the period of time that an object sent to Centera must be retained. This information is intrinsically associated with the object throughout its lifecycle. Once an object has been designated in this manner, the object cannot be deleted or overwritten before the expiration of the retention period. Furthermore, once the retention period has expired, Centera does not proactively delete expired content. Rather, deletion must be initiated by the controlling application.

Controlling applications communicate with Centera through a programmatic interface known as an Application Programmer’s Interface (API). The API allows only a predetermined set of Centera functions to be executed by the controlling application. Specifically, the API provides access to five basic functions, one of which is the “delete” command. However, the Centera software is written so that a “delete” command cannot be executed on an object that has an unexpired retention period. In this manner, Centera preempts the unauthorized deletion or overwriting of stored information.

## System 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 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.

**Centera Capabilities.** Centera offers a variety of controls and techniques designed to secure the system and its contents, as follows:

- 1) **Architecture.** Centera is not a “browseable” or directly-accessible system. The only access to Centera is through the controlling application or an administrative

WHERE LAW & TECHNOLOGY MEET

**KAHN**  
CONSULTING INC.

console. This architecture makes it difficult for an attacker without access to either of these entry points to find, view or access content within a Centera cluster.<sup>13</sup>

- 2) **Access Controls.** The controlling application's access to Centera can be policed in a variety of ways at the Centera System Administration level, including password protection and through configurable file operation (i.e., query, delete, retrieve, and store) protection. The type of access that applications have to virtual pools can also be controlled through an "access control list."
- 3) **Administration.** Centera can be configured to disallow remote administrative access. This would limit administrative access to individuals who have a direct physical connection to a Centera system, which typically would be located in a physically secure location. In addition, all management and configuration changes, as well as failed authentications, are logged and made available in a secure audit log.
- 4) **Application Access.** A controlling application's access to Centera is strictly limited to designated Internet Protocol (IP) addresses and port numbers.

## System Trustworthiness

**Desired Capability.** When organizations archive electronic records 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.

**Centera Capabilities.** The Content Address of a digital object is calculated *before* it is written to disk within Centera, and also immediately thereafter. Next, the two Content Addresses are automatically compared to detect any changes that may have occurred during the recording process. A background process also continuously runs to automatically recalculate and compare Content Addresses on a periodic basis. Additionally, Centera recalculates an object's Content Address when the object is read by the controlling application. These operations work to ensure that the object that is stored within Centera is the same object that was sent to Centera by the controlling application.

## Long-Term Retention and Access

**Desired Capability.** Systems designed for the long-term retention of electronic records should ensure that such records will be accessible for the period of time required.

**Information Management Principle.** Storing records in electronic form over the long term requires special attention to factors unique to the electronic environment that can threaten the long-term accessibility of electronic records. These factors include the limited lifespan of digital storage media; data corruption; heterogeneous storage formats; technological obsolescence; lack of access to hardware and software used to originally create the record; and so on.

**Centera Capability.** The health of each disk drive within Centera is continuously monitored. Disk and node failures that can be "self-healed" are repaired automatically by isolating the failed node and recovering its data onto other nodes in the Centera cluster, at which stage the failed node can be replaced by an administrator.<sup>14</sup> This capability effectively results in the ongoing migration of data from aged to fresh media. In addition, Centera's design is hardware-independent, which allows it to adapt to the latest storage technology. This important capability may also help to ease

WHERE LAW & TECHNOLOGY MEET

**KAHN**  
CONSULTING INC.

future upgrades or migrations to different media as new technology is developed and thereby help to ensure long-term access to stored electronic records.

## Business Continuity and Disaster Recovery

**Desired Capability.** Standard disaster recovery techniques require that data be stored in at least two physically separate locations. This is also a requirement of some regulations, such as SEC Rule 17a-4(f)(3)(iii), which requires that securities firms “[s]tore separately from the original, a duplicate copy of the record . . . for the time required.”

**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, destroyed, or otherwise made unavailable. Thus, there is a need for organizations to copy important data to different physical locations for backup, disaster recovery, and business continuity purposes.

**Centera Capabilities.** Centera software can be configured to continuously and asynchronously replicate the contents of one Centera installation or “cluster” to one or more physically separate Centera clusters. This capability will aid an organization in meeting its business continuity and disaster recovery needs as they relate to information stored within Centera. In addition, each Centera system can be powered by two independent sources of AC power, another capability that supports business continuity requirements.

## Destruction

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

**Information Management Principle.** Destruction is the final lifecycle stage of most information. In the digital world, it can be difficult and expensive to ensure that electronic information is properly destroyed. This can lead to situations where “deleted” files are recovered or recreated in the course of litigation, for example. In addition, the requirement to properly destroy of certain types of private information is a requirement of existing and emerging privacy laws and regulations in the US and abroad, including the Federal Trade Commission rules regarding the proper disposal of consumer information.<sup>15</sup>

**Centera Capabilities.** When Centera receives a request from a controlling application to delete an object, Centera first determines if the object has been assigned a retention period. If the retention period is still valid, then the object cannot be deleted. If the retention period has expired, the object is now eligible to be deleted. Consequently, upon receiving a delete command from the controlling application, Centera will delete the object, and automatically recover the disk space for further use.<sup>16</sup>

In addition, for added certainty in the destruction of objects, Centera can be configured to automatically use digital “shredding” techniques that conform to the US Department of Defense 5220.22-M (i.e., DoD 5015.2) standard for permanently deleting digital information.

It should be noted that a feature of Content Addressing is that identical objects are never stored more than once within Centera. For example, if the controlling application submits an e-mail message for archiving to Centera, and a calculation of that message’s CA reveals that it is identical to a message already stored within Centera, the message will not be archived again (which, among other things, promotes system storage efficiencies). Rather, a new Content Descriptor File (CDF) will be created that “points to” the original stored e-mail message, rather than storing an

WHERE LAW & TECHNOLOGY MEET

**KAHN**

CONSULTING INC.

additional, identical copy. Consequently, multiple CDFs can point to the same object, and the object would continue to exist until the CDF with the longest retention has expired and has been deleted. Among other things, this architecture supports situations where a single object is subject to multiple, different retention periods due to regulation or policy, or when the retention period for an object needs to be extended due to investigations, audits, or litigation.

Because Centera allows multiple CDFs with different retention periods to point to the same object, there will be cases where one individual no longer has access to a certain object because they have deleted “their” CDF, while another individual still has access to the object because “their” CDF still exists. When making representations to courts about the availability of information in the context of discovery proceedings, organizations should be aware of this feature of Centera and address its proper use through policy, procedure, and controlling application configuration.

## **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](http://www.KahnConsultingInc.com).

## **V. Endnotes**

<sup>1</sup> Pub. L. 107-204, 116 Stat. 745 (2002).

<sup>2</sup> “Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework,” Committee, Bank for International Settlements, November 2005.

<sup>3</sup> This Evaluation, completed in January, 2006, is an updated version of an Evaluation originally conducted in 2003.

<sup>4</sup> In undertaking this engagement, KCI exclusively relied upon information supplied by EMC through internal and external documentation, and interviews with EMC representatives, including senior system designers. KCI does not conduct independent laboratory testing of information technology products, and as such, did not evaluate Centera in a laboratory setting or otherwise field-test any EMC products.

<sup>5</sup> The threshold size can be configured by the administrator - the default setting is 250 KB.

<sup>6</sup> CPP requires a minimum of 8 nodes to function. As such, a Centera in a 4 node configuration would require two or more sets of 4-node Centeras within a cluster in order to offer CPP capabilities.

<sup>7</sup> FRE 901(b)(9).

<sup>8</sup> See, for example, *Kozlowski v. Sears Roebuck & Co.*, 73 F.R.D. 73 (D.Mass.1976).

<sup>9</sup> In the Matter of Banc of America Securities LLC, SEC Release No. 49386, March 10, 2004.

WHERE LAW & TECHNOLOGY MEET

**KAHN**  
CONSULTING INC.

<sup>10</sup> When a controlling application stores an object in Centera, it also creates and stores an Extensible Markup Language (XML) file containing the CA and metadata about the object. This XML file is known as a Content Descriptor File (CDF), and can contain both “standard” information such as filename and a time-date stamp, as well as “custom” metadata stipulated by the controlling application, such as a project name or office number, for example. The data in the CDF can subsequently be used for querying purposes, controlling retention periods, and for other purposes.

<sup>11</sup> Industry standard hashing algorithms such as those used by Centera operate in such a manner that the likelihood of two pieces of different information resulting in the same hash value is extremely low statistically.

<sup>12</sup> While a skilled attacker can circumvent even the strictest security controls within any information system, given enough knowledge, resources, and time, EMC had the Centera product reviewed by Internet Security Systems and @Stake, both of which concluded that the product was well protected.

<sup>13</sup> While a skilled attacker can circumvent even the strictest security controls and mechanisms within any information system, given enough knowledge, resources, and time, Centera has built in substantial features to prohibit such an attack and minimize any resulting harm to stored content.

<sup>14</sup> No single storage subsystem is immune from data loss if several hardware components fail simultaneously. EMC recommends that organizations employ the disaster recovery/replication features of Centera, which help to minimize the likelihood that data will be lost due to node failures or catastrophic events.

<sup>15</sup> “Disposal of Consumer Report Information and Records,” 16 CFR Part 682.

<sup>16</sup> It should be noted that a feature of Content Addressing is that identical objects are never stored more than once within Centera. For example, if the controlling application submits an e-mail message for archiving to Centera, and a calculation of that message’s CA reveals that it is identical to a message already stored within Centera, the message will not be archived again (which, among other things, promotes system storage efficiencies). Rather, a new CDF will be created that “points to” the original stored e-mail message, rather than storing an additional, identical copy. Consequently, multiple CDFs can point to the same object, and the object would continue to exist until the CDF with the longest retention has expired and has been deleted. Among other things, this architecture supports situations where a single object is subject to multiple, different retention periods due to regulation or policy, or when the retention period for an object needs to be extended due to investigations, audits, or litigation.

Because Centera allows multiple CDFs with different retention periods to point to the same object, there will be cases where one individual no longer has access to a certain object because they have deleted “their” CDF, while another individual still has access to the object because “their” CDF still exists. When making representations to courts about the availability of information in the context of discovery proceedings, organizations should be aware of this feature of Centera and address its proper use through policy, procedure, and controlling application configuration.

**Entire contents © 2005 Kahn Consulting, Inc. (“KCI”). Reproduction of this publication in any form without prior written permission is forbidden. KCI or EMC 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](http://www.KahnConsultingInc.com) [info@KahnConsultingInc.com](mailto:info@KahnConsultingInc.com) 847-266-0722**

WHERE LAW & TECHNOLOGY MEET

