



**Data  
Mobility  
Group**

[www.datamobilitygroup.com](http://www.datamobilitygroup.com)

**White Paper**

**12 June 2006**

*John S. Webster*



**Active Archiving**

**for Risk Mitigation *and* Business Reward**





## Executive Summary

Enterprise IT has always been ultimately charged with the protection and preservation of electronic records. Now IT is *legally* responsible for them as well. Stored information now represents a double-edged sword. It represents both business opportunity *and* legal liability. We believe that, as a result, managing legal and regulatory liability issues has taken precedence over aligning IT with business objectives and value creation. To regain a healthy balance between managing operational risk on the one hand, and aligning IT with business objectives and creating new business value on the other, IT can begin to relieve itself of the burdens imposed by regulatory compliance and litigation support by:

1. Implementing a tiered storage infrastructure (hardware plus software) that functions as an active archive.
2. Enlisting the help of records and information management (RIM) professionals who already possess the ability to implement records management policies.

A properly architected and implemented active archive can give both storage administrators and CIOs what they need as they move forward with information lifecycle management (ILM): automated efficiency for legal and regulatory records retention and retrieval coupled with indexing and search capabilities that can unlock business value. The well constructed active archive will also serve as an environment where IT and RIM professionals can converge and collaborate.

## A Double-edged Sword

During the last five years, two forces have combined to fundamentally change the ways IT managers store and retain electronic records. First, electronic corporate records including e-mail, have become the subject of legal discovery processes in support of lawsuits. Second, after a number of high profile examples of falsification and destruction of key corporate records including financial statements, royalty statements, and audit records, the US federal government under the Sarbanes-Oxley Act of 2002 created a legal obligation to store these records in such a way that they could not be modified, and required them to be made available to regulatory authorities upon request for up to seven years after they were created.



Business records have been traditionally (and correctly) viewed as strategic assets by at least one group of professionals – the Records and Information Management (RIM) community. They are accustomed to understanding the legal and regulatory ramifications that result from creating certain types of records. They also understand the business value of these records as strategic information, and have established standard and time-honored practices that help the enterprise unlock its stored business value.

However, understanding that there are now potential legal and regulatory liabilities attached to electronically generated records is relatively new to IT. These potential liabilities overlay electronic records with a significant amount of additional management complexity. Enterprise IT has always been ultimately charged with the protection and preservation of these business assets. Now IT is legally responsible for them as well. Stored information now represents a double-edged sword. It represents both business opportunity *and* legal liability.

DMG believes that, unfortunately for many IT organizations, managing legal and regulatory liability issues has taken precedence over aligning IT with business objectives and value creation. We observe that more resources are now being dedicated to regulatory compliance, exposure to legal discovery, and information security than to implementing new applications and creating new business processes that move the organization forward from a new business perspective. This sentiment was borne out for example in a 3Q05 survey of storage administrators by the SNIA's End User Council showing that speed of new application delivery received the lowest priority among six other operational responsibilities that included regulatory compliance, data protection, system reliability, and security.

There are at least two ways that IT can begin to relieve itself of the burdens imposed by regulatory compliance and litigation support:

1. Implement a tiered storage infrastructure (hardware plus software) that functions as an active archive. The active archive should have the ability to automate the placement *and timely retrieval* of information in such a way that legal/regulatory requirements are satisfied and that business value can also be continuously extracted from this long-term information source.
2. Enlist the help of RIM professionals who already possess the ability to implement records management policies in a physical records world. As an IT manager, you will find that they are now highly motivated to help you translate that professional knowledge to the world of automated electronic records management.



## Building the Active, Automated Information Archive

Building the active archive that functions as both a secure “vault” and a business information resource can be challenging when first approached. However, most if not all IT organizations already possess at least some of the basic building blocks. One of the first requirements is a pool of “secondary” disk storage that functions independently of the primary data store that is typically composed of high performance disk arrays. Many arrays now in production from the major disk array vendors also support SATA disk in addition to high performance disk. A secondary disk tier can be created by adding SATA disk trays to primary disk arrays already in service. A pool of secondary storage can also be carved out of older disk arrays currently in service but nearing the end of their productive use as primary disk. These arrays, often from multiple vendors, can be repurposed for archival service with the addition of a virtualized storage controller. However, the requisite data mobility and classification software must then be layered on top of the tiered storage infrastructure by the storage administrator.

Active archival solution sets that prepackage and integrate the basic requirements are also beginning to appear. We believe that to function as an active archival information resource for both risk mitigation and business enhancement, a solution set should offer:

- Straight-forward integration with the current IT application environment including the reuse/repurposing of existing IT resources
- Support existing and emerging electronic archive standards such as XAM
- Simultaneously support multiple application environments for both records retention and retrieval
- Demonstrate verifiable and legally defensible preservation of records in pristine form
- Scale to the petabyte range while preserving the required performance characteristics
- Tolerance for device failures without disruption
- Automated records management and retention policies
- Object-based capture of data, business related metadata as well as file and records classification metadata, and object interrelationships
- Automate the indexing of ingested objects
- Offer fast search and retrieval capabilities



- Operate across a number of different file types including electronic documents, email, and image files
- Support both on-and off-line storage resources such as tape and optical storage devices
- Data de-duplication to reduce the amount of storage capacity required as more and more data is retained before it is finally discarded

When these prerequisite functions are integrated within a single solution set, an IT administrator can then view it as the foundation for a long-term active archive that can meet legal, regulatory, and business requirements.

## Enlisting the RIM Professional

Building the active archive also represents an opportunity to establish a long overdue collaborative process between storage/IT professionals and the RIM professional community. The depth of RIM professional expertise was recently demonstrated at a Storage Networking World Panel entitled “Get the Real Story Behind ILM,” that consisted of both IT and RIM professionals. Here, one of the RIM panelists was asked who owns an electronic patient record in a healthcare setting – the patient or the care giver. This panelist was immediately able to quote federal guidelines and years of applicable case law relating to the subject.

RIM professionals can be valuable, time-saving resources for storage professionals when establishing and automating policy. A soon to be published paper coauthored by the SNIA and ARMA entitled “Collaboration – the New Standard of Excellence,” outlines the contributions RIM professionals can make:

A fundamental understanding of what electronic records really are – RIM professionals are grounded in best practices for records management – both physical and electronic. Many of these best practices can be translated to the IT management process.

Organizational records policies – Records policies monitored and managed by RIM professionals are in fact governed by processes that have been established over time by many other contributing professionals within the organization, including legal counsel and security staff.

Data retention schedules – There are over 900 regulations in the US alone that control records retention periods – electronic and otherwise. RIM professionals have an awareness of the breadth of the regulatory environment.



Record interrelationships – Separate and discrete records can be related to one another via many different circumstances and RIM professionals have a critical awareness of these interrelationships. Factors that must be considered include:

- The business reasons that caused the creation of a record set
- Security level
- Confidentiality of customer data
- Confidentiality of employee data
- Authorized usage
- Regulatory requirements
- Contractual agreements


An ideal active archive solution would be one that both storage administrators could use as a basis for collaboration. Therefore, the solution should be sensitive to pre-established organizational records policies and retention requirements, security requirements, and records interrelationships. It should also offer a user interface that can be approached by non-IT professionals.

## **Giving Storage Administrators and CIOs What They Need**

Researchers at the Information Storage Industry Center (ISIC) attached to the University of California, San Diego have released the preliminary findings of an ILM study. The study is aimed at reaching an understanding of what ILM means to IT users, from the storage administrator/operational level to the office of the CIO. It confirms some common understandings in the storage community – that ILM still carries much HSM baggage. At the same time it opens a window to some new ILM expectations that map directly to the active archive.

This study shows that, from data center operations to the CIO's office, the perceptions of the value of ILM change. For storage administrators, the value of ILM is bound up in wringing greater efficiency out of a networked - and now tiered - storage infrastructure. The CIO's perception of ILM is dramatically different. Here, the expected value of ILM is expressed in terms of delivering business intelligence to decision makers.



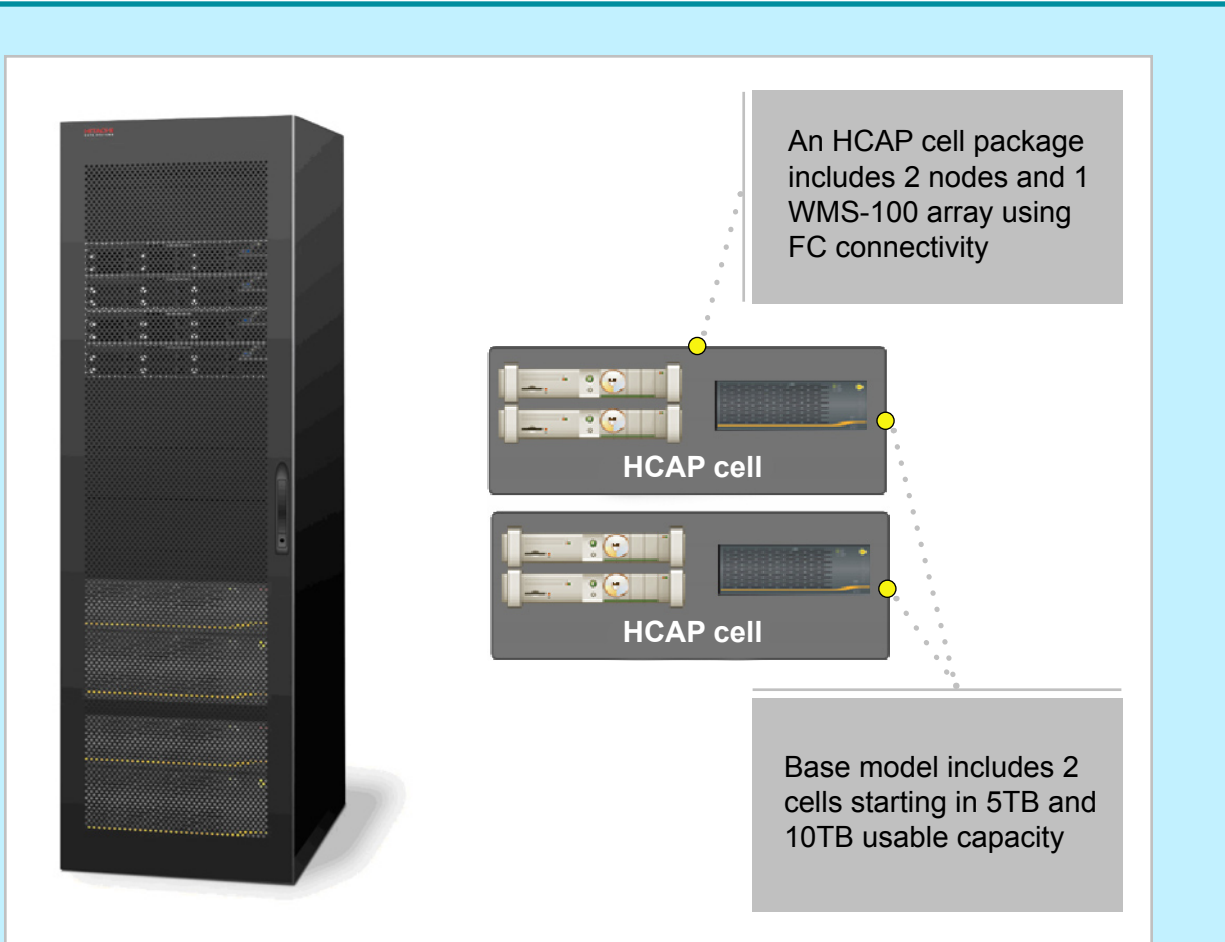
A properly architected and implemented active archive can give both storage administrators and CIO's what they need as they move forward with ILM: automated efficiency for legal and regulatory records retention and retrieval coupled with indexing and search capabilities that can unlock business value. The well constructed active archive will also serve as an environment where IT and RIM professionals can converge and collaborate. 

## for example

### **Hitachi Data Systems' Content Archive Platform**

On June 12, 2006, Hitachi Data Systems extended its tiered storage offerings to include an active archival storage solution as part of its Application Optimized Storage™ (AOS) portfolio. Called the Hitachi Content Archive Platform, it is an integrated solution set consisting of hardware, software and services aimed at helping its customers establish an active archival storage environment in support of enterprise ILM objectives. Using the Hitachi Content Archive Platform framework, Hitachi Data Systems' customers can construct a storage environment that provides for the long-term preservation and protection of immutable content that is also made available and easily accessible for retrieval in a timely manner for both risk mitigation and business intelligence needs.

Hitachi Content Archive Platform is built on a RAID + SAN in an array of independent node (SAIN) architecture model.. The Hitachi Content Archive Platform is delivered as a pre-configured, pre-integrated package that consists of servers, software, and Hitachi Data Systems TagmaStore® Workgroup Modular Storage (WMS) storage systems, and is available in 5 and 10 terabyte configurations. Hitachi's line of intelligent virtual storage controllers, the TagmaStore Universal Storage Platform (USP) and Network Storage Controller (NSC) will be able to work with the Content Archive Platform to virtualize the content archive tier. The initial, base configuration consists of a four server-node cluster that is scalable in two-node increments that is pre-packaged and pre-configured for customers.



**Key Hitachi Content Archive Platform attributes include:**

Multi-Application Access — Built-in NFS, CIFS/SMB, HTTP and WebDAV gateways provide standards-based archival storage access by multiple client applications. Fixed content files including email or data and descriptive metadata are ingested from producing applications. Files are stored in native form with original names.

Object-based Archive Data Management — Archived objects consist of data, metadata, and associated policies, but are not bound to specific applications. Objects have their own policies or inherit from parent directories or file systems. The Hitachi Content Archive Platform also supports large fixed-content files such as images, voice recordings, geospatial data, etc.



Distributed Architecture — Each Hitachi Content Archive node serves both as an ingest point for archived content and a storage node for storing archived content that can be configured to support remote archive sites.

Search & Indexing — Metadata provides a searchable index to access and retrieve content. The Hitachi Content Archive Platform provides a web interface that can be used by both IT and non-IT users to retrieve documents by searching keywords within the file. Data can also be accessed like a standard file system for browsing and provides drag and drop functionality.

Self-Managing — The Hitachi Content Archive Platform is self-configuring, self-healing, and simplifies the process of increasing capacity to meet demand.

Linear Scalability — The Hitachi Content Archive Platform scales to petabytes with a single file system using a symmetric file system model and underlying SAN architecture, and scales to 350M objects within a single cluster.

Automated Policy — Policies can be automated that manage object retention periods, authentication (digital signature), protection, and export to other media

The Hitachi Content Archive Platform also leverages OAIS model and, as such, puts it in advantageous position with regard to IT/RIM convergence and support from content management vendors. Hitachi Data Systems also supports the SNIA's XAM initiative.

During the first half of 2007, Hitachi Data Systems also plans to provide storage virtualization support for the Hitachi Content Archive Platform. At that time, an intelligent virtual storage controller (Hitachi TagmaStore USP or NSC) can be inserted in front of an existing Hitachi Content Archive Platform or in front of older arrays repurposed for archival usage to perform a number of functions essential to the active archive including:

- Serving as a platform for data mobility processes i.e. the movement of data from primary to secondary disk and tertiary tape storage tiers, in effect acting as a “gateway” between primary high performance storage and archival storage.



- Serving as a platform for other essential archival storage services such as classification.
- Mask the differences and achieve interoperability among heterogeneous arrays that are part of the archival storage pool.

The Hitachi Content Archive Platform is not content addressable storage (CAS) and Hitachi Data Systems has not tried to mimic its competitors while creating its Content Archive Platform. Rather, it created a solution set that is highly scalable and is more in-line with the convergence and needed collaboration between IT and RIM professionals. It addresses the needs of storage and IT administrators for compliance and legal discovery, but offers the CIO an ability to leverage the active archive for use in BI types of applications as well.

Copyright© 2002-2006 Data Mobility Group, LLC. All Rights Reserved. Reproduction of this publication without prior written permission is forbidden. Data Mobility Group believes the statements contained herein are based on accurate and reliable information. However, because information is provided to Data Mobility Group from various sources, we cannot warrant that this publication is complete and error-free. Data Mobility Group disclaims all implied warranties, including warranties of merchantability or fitness for a particular purpose. Data Mobility Group shall have no liability for any direct, incidental, special or consequential damages or lost profits. The opinions expressed herein are subject to change without notice.

This report sponsored by Hitachi Data Systems. All other brands, products, or service names are or may be trademarks, or service marks of, and are used to identify, products or services of their respective owners.