WHITE PAPER
April 23, 2004

Strategic archiving

Using information lifecycle management to archive data more efficiently and comply with new regulations

ABSTRACT

Information lifecycle management can help an organization address the archiving challenges brought by new regulatory requirements and runaway data growth. The use of a tiered approach to archiving allows an organization to leverage existing assets to store data more efficiently and at a lower cost while meeting external and internal guidelines for data retention. Information lifecycle management can be applied today to address the needs of a wide range of application areas, including healthcare, financial services, broadcast and e-mail.

1 Executive summary

- With evolving uses of data and changing regulations, archiving is no longer just a back-end process. It is now also a front-end process focused on data accessibility.
- A total systems approach to archiving can help an organization optimize the use of existing assets and storage tiers to archive data in a manner that increases efficiency and compliance.
- While specific application requirements vary, a common approach to archiving can be implemented today in a three-step process: assess information assets, adapt the infrastructure, maintain balance over time.
- The application of more strategic approaches to archiving has helped organizations regain lost productivity, increase compliance with data retention requirements and enhance access to archived information.

2 The challenge

2.1 Taking control of archiving and compliance

Historically, archiving has too often been treated as an afterthought. Archiving is misconstrued as what happens to data when people are done using it but not ready to throw it out. Cast into this role, archiving has stood in the shadows of higher-profile storage activities, such as those focused on increasing the availability of primary data stores and ensuring business continuity. But today this is rapidly changing. The twin challenges of fast data growth and increasing regulatory retention requirements are forcing enterprises to take a close look at their data retention policies and archiving tactics.

No longer is archiving simply a back-end process that involves finding an inexpensive place to leave data for a long time. It is now also a front-end process focused on making archived data readily accessible to those who need it — whether it's a compliance officer responding to an inquiry from a regulator, a surgeon who needs immediate access to a five-year-old cardiac imaging study or a broadcast producer who wants to repurpose an aging news clip.

Across a wide range of industries, organizations are seeking advanced archiving solutions that enable them to cost-effectively retain and retrieve large volumes of data as required to achieve business goals. In certain industries, those goals include compliance with new regulatory requirements from external agencies and new legal guidelines from internal management. In a study by the IDC research firm, 51 percent of surveyed companies indicated that compliance issues have strong or moderately strong influence on their data protection and archiving strategies.[1]

Heavily regulated industries, such as healthcare and financial services, face stringent requirements for data retention. In the United States, these regulations include the Health Insurance Portability and Accountability Act (HIPAA), regulating patient medical data, and the, Securities and Exchange Commission regulations (17CFR240.17a-4) regulating business and financial record keeping.

[1] *IDC study, October 2003, 235 respondents.*

Other industries, including broadcast and video surveillance, need to find ways to economically archive ever-growing numbers of large digital data files. And virtually all enterprises wrestle with massive amounts of e-mail generated in the course of day-to-day business. In general, electronic business information, even information created in rank-and-file client systems, can no longer be considered unofficial or disposable. Enterprises everywhere are challenged with the need to retain more of their information for longer periods and to keep that information readily accessible to users.

E-mail, in particular, is a formidable challenge from both a management and a legal point of view. For example, a recent study of corporate e-mail users showed that e-mail volume per user is growing by 20 percent annually2, while the number of corporate mailboxes is growing by 16 percent per year[2]

2.2 Cross-industry challenges

While specific data retention requirements and applications vary by industry and country, the underlying management and infrastructure issues tend to cut across industry lines. Businesses everywhere are dealing with a common set of archiving issues. Among them:

- Archiving costs are increasing.
- Data volumes are growing faster than IT budgets.
- Historical data is consuming expensive primary storage.
- Legal liability for inappropriately retained or deleted data is growing.
- Older data is trapped in obsolete platforms.
- Retention and deletion policies are not consistently employed.
- Some data retention policies require management capabilities that are not in place.

These are among the challenges that can be addressed by applying the principles of information lifecycle management to an archiving environment. This holistic approach to information management provides a framework for more strategic archiving.

3 Information lifecycle management

3.1 Employing a tiered storage hierarchy

Information lifecycle management is based on the idea of storing and managing data according to its value and purpose. It recognizes the inherent differences in the value and use of data sets. It makes use of tiers of storage to match archiving options with the appropriate cost, security, performance and accessibility requirements for particular data sets. Information lifecycle management takes a total systems approach that considers the full range of storage options — from disk to tape to optical.

[2] *Radacati study, E-mail Archiving Market Needs, May 2003.*

This holistic approach avoids the use of a single technology to archive diverse types of data. The efficiency in information lifecycle management lies in the ability to use different tiers of storage to meet the individual requirements of a particular data set. The use of a tiered archive methodology helps an organization lower the total cost per terabyte of archived data while complying with information retention and protection regulations. Lower-cost storage tiers help hold the line on storage costs. Higher-performance storage tiers provide faster access to archived data that might suddenly become extremely important to the business.

By applying information lifecycle management, organizations can improve the efficiency and scalability of their data archives to meet changing compliance and business requirements. This more strategic approach to archiving enables new data-intensive business opportunities and helps minimize regulatory liability.

3.2 Data classification and automatic policy management

Regulatory requirements are increasingly driving different needs for different data types. Under SEC regulation 17a-4, broker dealers must save stock sales information for not less than six years, yet they must maintain customer account information for not less than six years after the closing of any customers account[3]. Data archiving needs vary by the type of information being retained and the applicable compliance regulations.

Data classification is one means of providing support for these varying requirements. Classifying data and automating policy management can maintain data for the required periods and not a moment longer (if desired). IT administrators can set rules for retention based on standard data classification schema so data automatically goes into the proper retention framework.

Retention is not the only requirement driven for regulatory structures. For example:

- Security of patient data is critical to healthcare providers under HIPAA regulations. Patient data must be securely handled. Dissemination to unauthorized parties is strictly forbidden under HIPAA. This may require the electronic transmission of such data to be encrypted. Data classification and automated policy management may play a crucial role here in establishing a secure environment for patient data whenever it is transmitted within the corporate infrastructure.
- In some government environments, secure deletion is a requirement. This involves a government approved process of "multiple overwrites" of secured data to insure no unauthorized access to deleted data is possible.
- Some regulations require the ability to audit data activity to insure compliance. Archiving systems may have to be able to produce a record of every data transaction that could affect the data.

3 "Positioning Storage to Support Information Life-Cycle Management in a Regulated World." Bob Zimmerman. Giga Research. December 19, 2003.

In countries other than the United States, the regulatory landscape is similar. Companies must be ready to account for any applicable regulations wherever their business takes them. Data classification and automated policy management can ease the burden of managing multiple regulatory requirements.

3.3 Deleting data as appropriate

While meeting data retention and availability requirements, organizations must also maintain appropriate policies for the deletion of data. Without policies for eliminating data that has reached the end of its lifecycle, an organization and its IT resources could eventually be overwhelmed by useless data.

Yet for many companies, data retention policies are rarely used. Where such policies do exist, archiving and data management systems are often inadequate or simply not in place. As a result, archived data may simply pile up year after year, driving up an organization's cost of storage.

Every organization should have clear-cut policies on the retention and deletion of all categories of stored information. Giga Research advises that every application should apply a lifecycle policy to every data element that supports critical business decisions.[4]

Information lifecycle management isn't just about data preservation; it is also about data deletion. Under this more strategic approach to archiving, data should not be allowed to simply pile up indiscriminately on disk, tape or any other storage option.

3.4 Leveraging current assets

A thoughtful approach to information lifecycle management does not hinge on the availability of new technologies, nor does it require the purchase of costly hardware or software solutions. Instead, ILM focuses on making more efficient and complete use of existing storage assets.

This might mean automating the process of migrating older data from primary disk to automated tape systems. This can free up valuable primary disk space while making use of underutilized tape assets.

Many organizations have a great deal of unused storage capacity in tape libraries that are attached to mainframe systems. With the addition of Fibre Channel tape drives, this storage asset could be made available for Open Systems archiving. Similarly, the consolidation of data center assets can produce archiving economies driven by the use of fewer, better-utilized storage resources.

These are just a few examples of how an organization might implement an information lifecycle management strategy without making major investments in hardware and software. With ILM, an organization can build on its existing archiving infrastructure without starting over.

[4] *"Positioning Storage to Support Information Life-Cycle Management in a Regulated World." Bob Zimmerman. Giga Research. December 19, 2003.*

4 Applying the ILM strategy

4.1 A three-step process

A total systems approach to information lifecycle management can be put to work today across a wide spectrum of applications to improve archiving productivity and compliance. This strategic approach to archiving can be applied in a three-step process:

- Assess information assets and uses.
- Adapt the storage infrastructure to store information according to its changing purpose and value.
- Maintain the data balance with integrated information management tools.

4.1.1 Step one: Assess

Improving archiving for compliance begins with an assessment of access requirements and retention regulations for different data types and ages. Data should be stored to match these varying requirements and maintained with automated management and migration tools.

This first step in the assessment process strives to identify the current state of the environment. Questions addressed here include:

Matching data retention requirements with archiving options

	Higher cost			Lower cost
Higher value, higher recall ability	Primary disk		S-ATA disk	
Lower value, lower recall ability		WORM disk storage		WORM tape and optical storage

- What is the issue? What parts of the business have the potential for greatest impact or risk?
- What regulations apply to these businesses?
- What is the risk associated with noncompliance?
- Is more frequently accessed data stored on higher-performing systems or media?
- Is rarely accessed data placed on the most cost-efficient media?

The likelihood that data will need to be recalled is a key consideration in archiving policy decisions. In general, the less likely that data will need to be recalled, the stronger the argument for moving it to a long-term archiving system. Giga Research advises that if more than 1 percent of files stored in long-term archives are recalled, chances are that data is being archived too soon.[5]

[5] *"Positioning Storage to Support Information Life-Cycle Management in a Regulated World." Bob Zimmerman. Giga Research. December 19, 2003.*

With a total systems approach to archiving systems, data is retained based on its likelihood of recall, using a tiered archive methodology. This approach strives to lower the total cost per terabyte of archived data — while complying with information retention and protection regulations for its recall ability.

A scheme for matching data with appropriate archiving options might take the form of the following table.

Ultimately, the assessment process should yield the information necessary to design a more strategic archiving solution. The design process should result in these deliverables:

- A strategy for dealing with the issue.
- Options for many levels of archiving.
- Options for many levels of compliance readiness.
- The key components of people, process and technology to deploy the solution.

4.1.2 Step two: Adapt

The following step in the ILM implementation process involves targeted actions to adapt the archiving infrastructure to address identified issues. Actions here include:

- Deploy technologies to build or help reinforce archive and compliance.
- Document processes to maintain proper archive and compliance.

The solution deployed in this phase should improve archiving efficiency and compliance with both external regulations and internal policies. There is no such thing as a dead-end archive. All archived information should be able to be recalled to meet compliance and business needs. Compliant systems should meet requirements for information integrity, accessibility and retention periods.

Automation is key here. Data should be stored and maintained to match varying requirements with automated management and migration tools. With a total system approach to archiving, archiving capacity can grow with much less cost because only a fraction of the archived data need be stored on higher-cost disk.

4.1.3 Step three: Maintain

The third step in the process of implementing information lifecycle management focuses on maintaining the data balance. Actions in this ongoing process include:

- Perform regular audits to monitor archiving and compliance standards.
- Promote customer satisfaction by working to meet service level agreements.

The work here uses integrated information management tools to automate the adjustments to the storage infrastructure. The goal is to keep data performance, protection and retention systems in line with data's changing role. Storage resource management tools and well-defined change management processes can keep the information infrastructure optimized through normal business change.

Maintaining the balance also includes ongoing examination of the classifications used to match data sets with the ideal archiving options. What percent of data is currently classified as critical? Has the data of certain users risen in importance to the business? Do access and performance requirements need to change to address shifts in business priorities? These are the kinds of questions that should be reviewed on a periodic basis.

This ongoing work enables the realization of the information lifecycle management vision — in which data is archived according to its value to the organization.

5 The opportunity

A total systems approach to addressing information archiving requirements can generate significant, quantifiable gains. These gains stem from the implementation of systems and processes that enable more cost-effective approaches to information management, primary storage, data protection, and archiving and compliance.

The opportunities presented by a total systems approach to archiving are illustrated in the following case studies.

5.1 Healthcare applications

The creation of a consolidated digital archive helped a hospital system reduce its costs of storing patient imaging studies, improve data access times and position itself for easier compliance with federal patient-privacy regulations. The tiered storage solution makes use of different levels of disk and tape to archive imaging studies in a cost-effective manner.

When data is created, it is stored in a disk cache and simultaneously written to high-speed tape cartridges in an automated library. When it is three weeks old, data is automatically deleted from the disk cache to free up space for new imaging studies. The data remains on the high-speed tape for two years to give the medical staff fast online access to archived files. After two years, data is migrated to a high capacity tape cartridges for low-cost long-term storage.

5.2 E-mail applications

A global technology company used information lifecycle management principles to address skyrocketing e-mail growth. The e-mail generated by more than 5,000 employees was growing faster than IT resources. This situation resulted in restrictive policies that reduced end-user productivity.

An information lifecycle management assessment identified the opportunity to use policy-based data management to create what amounted to bottomless mailboxes for users. The ensuing implementation of a disk, tape and software solution led to dramatic gains. The company recovered 321,000 hours of wasted end-user productivity, reduced total cost per e-mail user by 75 percent and recovered nearly one terabyte of network storage.

5.3 Financial services applications

A growing financial company achieved substantial gains by deploying a payment processing and archiving solution that combined scalable online disk with lower-cost media and software intelligence for automated information movement.

The solution resulted in a 630 percent increase in the speed of storing information to disk, a 99.5 percent decrease in data access time and reductions in backup windows.

5.4 Broadcast applications

Creation of a digital video archiving system helped a national television station reduce its cost of storing broadcast files, automate labor-intensive management work and streamline the process of sharing news content with subordinate channels.

The solution leverages an automated tape library, high-speed and high-capacity tape drives, and application-specific software and databases to provide complete automated data management, access and query functions.

6 Recommended actions

- Effective information lifecycle management requires tiers of storage, so consider the full range of archiving options, from disk to tape to optical. Avoid single-technology approaches to archiving diverse types of data.
- Before making changes, thoroughly assess your current archiving issues and assets. Identify areas of greatest need in terms of regulatory compliance and management challenges. Set priorities accordingly.
- Involve a wide range of stakeholders in archiving and compliance decisions. This might include IT staff, corporate compliance officers, human resources staff, risk-management teams, and representatives of auditing, finance and legal departments.

7 Planning considerations

- Don't take a short-term view. Think in terms of 10, 20 or 30 years. Archiving and compliance requires long-term strategies that encompass migrations to newer, faster technologies.
- Recognize that compliance is an ongoing process. Develop your plans with the assumption that regulations will continue to change.
- Remember that technology is only part of the answer. It's also important to examine the processes used with technology to maintain control of your data and its uses.

ABOUT STORAGETEK®

Storage Technology Corporation (NYSE: STK) is a $2 billion global company that enables businesses, through its information lifecycle management strategy, to align the cost of storage with the value of information. The company's innovative storage solutions manage the complexity and growth of information, lower costs, improve efficiency and protect investments. For more information, visit www.storagetek.com, or call 1.800.275.4785 or 01.303.673.2800.

WORLD HEADQUARTERS

Storage Technology Corporation
One StorageTek Drive
Louisville, Colorado 80028 USA
1.800.877.9220 or 01.303.673.5151

TT 0013 A 06/05