ESg
Enterprise Strategy Group

IT Knowledge • Business Results

Market Review

Building An Intelligent Active Archive -

Transforming Passive Data Retention into

a Reliable Tool for Information Governance

By Peter A. Gerr, Senior Analyst
Data Management and Data Protection Solutions

May 2005

Table of Contents

Introduction

Information is the currency of today's global, networked economy and a number of current business realities – including business continuity concerns, regulatory compliance, information security threats, and continual growth in data – are combining to place a greater premium than ever on effective enterprise data protection and preservation policies. With more than 90% of all global corporate information being created in electronic or digital format[1] companies spend billions of dollars annually on mature and emerging technology solutions that hold the promise of improving core data protection processes like backup and recovery. Yet failure rates related to these most basic processes remain high, as do users' concerns that at least some of their data is unprotected. While the explosion in digital data has been researched and forecasted by dozens of academic and analyst firms, it seems as if all attempts to size the amount of information being created fall short of capturing the true impact to both IT leaders and their business counterparts. Furthermore, while the rapid growth in information creation puts more strain on IT staffs and budgets, the bigger challenge is that more of this information must be retained for longer periods of time while still remaining accessible for rapid recovery.

The byproduct of information's emergence as the new global currency means that, like physical assets such as servers, networking equipment, and, ironically, the storage systems on which data is stored, information must be protected and preserved throughout its lifecycle. Increasingly, business and IT leaders are realizing that good corporate governance is not simply a best practice, but can also be a competitive differentiator when applied to certain information assets. Yet without also addressing how corporate intellectual property (IP), in the form of electronic records, spreadsheets, and other data, is managed and governed through its lifecycle, any corporate governance initiative is incomplete.

Capacity of Reference Information Worldwide, 2001-2006
Worldwide, Digitized Reference Information created & stored will surpass Non-Reference Information by the end of 2006
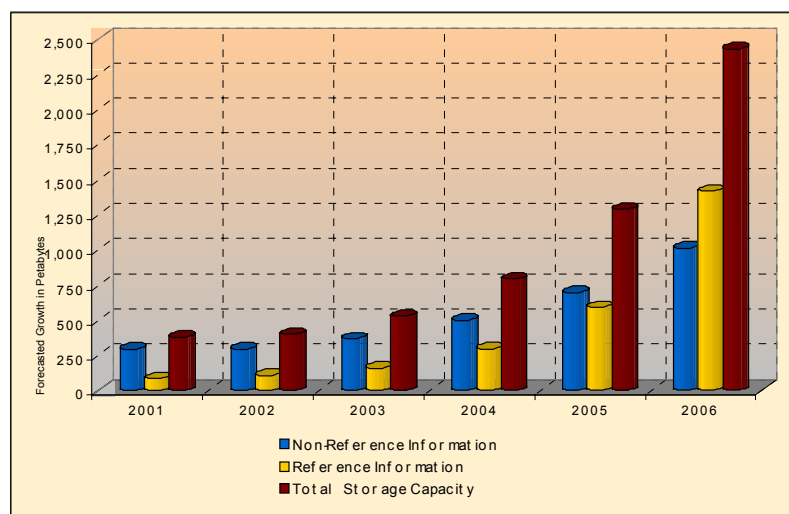
Source: Enterprise Strategy Group, 2002

Meanwhile, the business-level impact of poor data protection and preservation practices, which leads to ineffective preservation of corporate IP, continues to gain momentum and visibility. The fact that US and international courts now consider electronic files (email and email attachments, spreadsheets, documents, medical images, etc.) to be the equivalent of paper records and therefore a legal history of business decisions has been a boon for prosecutors and regulatory auditors who now rely heavily on the content and context of electronic communications for critical legal evidence. Many of the scandals that have rocked publicly traded companies and Wall Street financial firms share these common elements - the use and misuse of electronic communications, corporate, and personal data and the increasing relevance of electronic records as evidence in prosecuting and defending these cases. Whether a particular incident of corporate malfeasance results in actual litigation, the damage is done - customers, business partners, and the broad financial markets perceive any such loss, misuse, or theft of information as negligent, and an indictment,

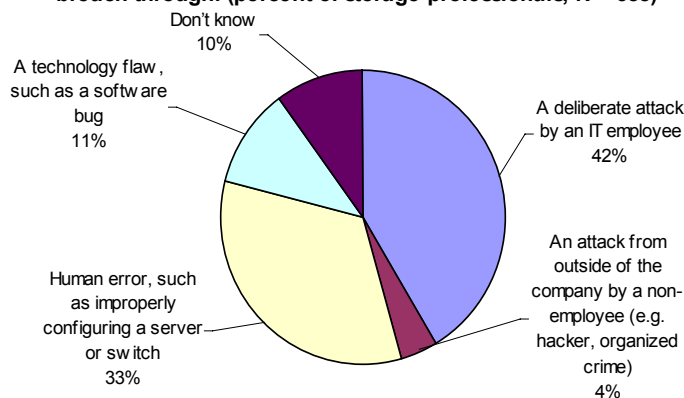[1] *Legal Times*, "What you can't see can hurt you", 09/27/2004

at least, of that company's general business practices. The following list contains illustrative examples of world-class organizations that have failed to make information governance a core competency, and were adversely effected by resulting litigation or damage to its reputation. Additionally, several incidents in the recent past involving the apparent loss or theft of backup tapes that contained corporate data further illuminate the considerable risks facing businesses today:

- 2001 - <u>Enron Corp. and Arthur Andersen LLP</u> – Andersen put in practice a document retention policy that called for destroying unneeded files just as the SEC began looking into Enron's finances. Andersen was forced to surrender its accounting license after it was convicted of destroying documents (its appeal is being reviewed by the Supreme Court)[2] and was later required to pay more than $60 million to settle civil lawsuits.
- 2004 - <u>Phillip Morris USA Inc.</u> – sanctioned $2.75 million after it was ruled that 11 employees failed to preserve emails, as was required by the company's retention policy and the court's preservation order.[3]
- 2005 - <u>Morgan Stanley</u> – found "grossly negligent" by a Florida state judge in failing to produce emails and other documents related to Sunbeam Corp. in response to financier Ronald Perelman's $2.7 billion lawsuit.[4]
- 2005 - <u>Bank of America</u> - admits that it has lost backup tapes that contained personal information including Social Security Numbers and account information for some 1.2 million federal employees. At the time of this paper's publishing, this incident was still under investigation.[5]
- 2005 - <u>Time Warner Inc.</u> - reports that some 40 tapes containing Social Security numbers and other personal information about 600,000 current and former employees, some of their dependents, and beneficiaries have been lost, apparently by a 3rd party services company. At the time of this paper's publishing, this incident was still under investigation.[6]

These real world examples demonstrate that failure to manage and ultimately mitigate the risks associated with unlawful and inappropriate use of digital data can irreparably damage a company's reputation and, by extension, shareholder value. A more troubling consideration is that the companies mentioned above spend tens of millions of dollars annually on technologies to protect, share, and maintain data and corporate IP. Yet each has been negatively impacted for its inability to appropriately preserve and archive business records – a convincing indictment of the procedures these firms followed, or didn't follow, to ensure such data was preserved. Conversely, these examples represent a failure of technology itself, or more

Threats to Information Security are Diverse

I believe that my company is most vulnerable to a storage security breach through: (percent of storage professionals, N = 388)

- Don't know 10%
- A technology flaw, such as a software bug 11%
- A deliberate attack by an IT employee 42%
- An attack from outside of the company by a non-employee (e.g. hacker, organized crime) 4%
- Human error, such as improperly configuring a server or switch 33%

Source: Enterprise Strategy Group, 2004

---

2 Forbes.com, 01/07/2005
3 Shook Hardy & Bacon, 2004
4 The Wall Street Journal, page C3, 03/22/2005
5 The Associated Press, 03/01/2005
6 CNN, 05/02/2005

specifically, the storage and archival solutions these firms relied upon to protect valuable information from such losses – in other words, the archive, in this case, failed in performing one its fundamental jobs.

Even more recent attempts by businesses to address the impact that compliance and records management regulations like Sarbanes-Oxley, SEC Rule 17a-3, and HIPAA have on the archival and lifecycle management accommodate only a subset of the requirements that enterprises must now consider. The pervasive use of and reliance on electronic information in today's business climate magnify the existing operational, legal, and regulatory challenges companies and their leaders face, yet are only a few of the threats that executives must now consider as they develop and deploy an archival strategy that enables sound information lifecycle management (ILM).

As companies inevitably continue to create, share, and retain greater volumes of business-critical, valuable, or volatile information assets, the severity and impact of such threats will increase accordingly. While public scandals such as the ones mentioned above underscore the importance of addressing these threats proactively, a surprisingly low percentage of companies are actually doing so. According to a recent survey of global CEOs by PriceWaterhouseCoopers, only 1 in 4 global CEOs believe that they are managing governance, risk management, and compliance initiatives effectively.[7] The reality that business leaders face today is that they are no longer simply responsible for the livelihood of the companies they lead – business, legal, and technology executives must all recognize their new roles as part-time stewards of the digital mountains of information created, shared, disposed of, or, more likely, retained for some future purpose. **Regulatory compliance, the threat of data loss or theft, the unrelenting explosive growth in electronic data, and the ongoing challenge to reduce IT expenditures represent but a few of the many drivers that are fundamentally altering the traditional definition of *archiving*. These potent challenges are collectively forcing IT and business professionals alike, to consider what combination of technology solutions will satisfy regulators, but more importantly, allow the business to develop and maintain an intelligent online or active archive of all digital or digitized reference data assets.**
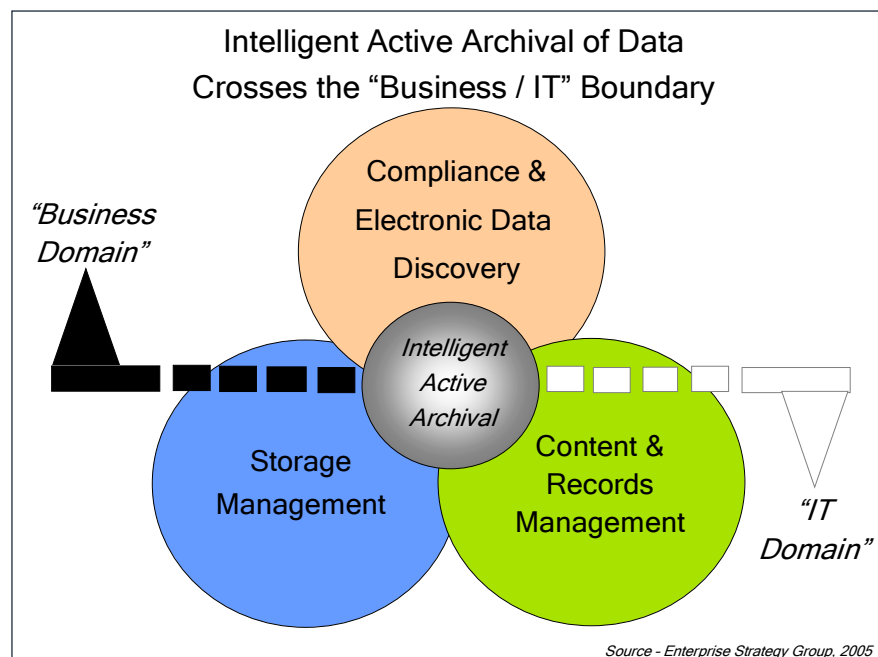
---

7 8th Annual PriceWaterhouseCoopers Global CEO Survey, 2005

The Liabilities of Today's Archive Solutions

The primary objective of building an archive is to preserve records that have value and relevance to the business or perhaps those that are required to be retained in an unalterable form to meet a compliance or governance mandate. For global corporations or large enterprises that have IT budgets measured in the tens-of-millions of dollars, this objective may, at first, appear to be straightforward. Individually, each of the aforementioned challenges pose a threat, yet are still merely symptoms of a larger issue - which is how organizations will manage not only the Terabytes (TB) of new information being created annually, but the Petabytes (PB) and eventually, Exabytes (EB) of information, representing billions of "data objects" that collectively make up an organization's history, and as such, must be retained.

This "perfect storm" of technology challenges, business threats, and regulatory mandates is changing the definition and role of archiving, and forcing users to reevaluate their data protection and preservation processes as well as the underlying technologies that enable accountability. Companies need to proactively protect themselves from ongoing misuse of communications channels, but also prepare their organizations for legal or compliance-related events that require fast and effective discovery of potential evidence. While more emphasis than ever is being placed on sound records management policies and governance of corporate information, we're really just beginning to address the long-term management of electronic records.

An Intelligent Archive can Transform a Captive Liability into an Accessible Asset

Intelligent Active Archival of Data
Crosses the "Business / IT" Boundary

"Business Domain"

Compliance & Electronic Data Discovery

Intelligent Active Archival

Storage Management

Content & Records Management

"IT Domain"

Source - Enterprise Strategy Group, 2005

While archiving, that is, retaining data on removable media such as tape or optical in a separate physical location, has long been a standard practice, the rules, as they say, are clearly changing. Previous generations of archives were built primarily as long-term repositories for all corporate data without consideration given to the actual value of or potential future uses for that data. Typically, earlier archives were extensions of IT's backup processes, and in many cases the archival process itself consisted of shipping backup tapes or optical platters to an offsite facility for long-term retention. While effective, perhaps, as a means for protection against a catastrophic systems failure or other disaster that resulted in the loss of primary data, such an archive provides minimal value beyond this. Without online or active access to historical data these assets are, in a sense, captive and unable to be used or reused for a future purpose.

The decade-long decline in the cost of disk storage coupled with the more recent emergence of new types of business-class disk media (P-ATA, S-ATA, SAS) enables more data to be retained online in a more readily accessible format, which certainly helps to address the accessibility part of the problem. Especially when an organization must quickly recover data, perhaps for a deadline-driven litigation or electronic discovery event, the ability to quickly search and retrieve only the electronic records pertinent to a case or event is clearly an advantage over previous archives. Yet while moving a corporate archive "online" by leveraging disk technology

solves the problem of needing timely access to information, blindly retaining historical records of all types in this manner in no way satisfies the contemporary compliance and governance requirements of today. While building an online or active archive is a step in the right direction, many existing disk-based solutions are fraught with limitations – without intelligence to securely capture, manage and protect, and ultimately enable valuable information assets to be repurposed or reused, a disk-based archive simply becomes an electronic dumpster, leaving IT administrators unable to differentiate between what records have value (and should be archived) and which have outlived their usefulness (and should be purged).

A truly intelligent active archive, therefore, not only provides the requisite features to enable IT to simply and effectively manage data assets appropriately, but also crosses the boundary to become an asset to the business. By providing a consistent means to capture, preserve, and most importantly, provide access to all corporate data via an easily searchable interface, the intelligent active archive actually becomes more valuable to the business at time passes. Although the appropriate levels of security, access, and control must be enforced by the intelligent archive, individual users and line-of-business managers can now have a rich, complete, and accurate historical view of all corporate information assets – to build new applications or customer-facing services, to offer a higher standard of customers support, or to enable corporate counsel to quickly and confidently locate the records needed to defend the organization from litigation – activities which simply are not possible with a static, offline, or unintelligent archive.

Furthermore, while there is clearly a move by companies to replace or augment their existing tape infrastructures with disk-based solutions, these same users rightfully value the portability and cost-effectiveness of tape especially for disaster recovery and business continuance purposes. **The intelligent active archive will balance the use of storage medium as appropriate to address the use and value of the data itself, in other words, retaining data on disk when rapid access is a priority, and migrating the data to tape or other removable media to enable cost-effective preservation with massive, but offline, scalability. The bottom line is that incumbent approaches to archiving must evolve and so to must the technologies that companies have at their disposal to proactively address the increase in the volume, value, and volatility of corporate information assets.**

Basic Challenges of Data Management & Data Protection

The new and historical threats facing corporate data today are magnified by IT's perpetual struggle to perfect the basics of data management, protection, and by extension, preservation. Some of these pitfalls are known and therefore manageable, while others require constant and proactive protection. Still others lie directly in the path of an organization's ability to confidently retain, preserve, and retrieve all relevant business records, for example:

- o Data Grows while IT Budgets and Staff Decline - Unrelenting growth of data, combined with flat or declining IT budgets and/or staff, creates an ever-widening gap between what is possible for IT to undertake and the service levels IT is able to support. IT staff is consumed with reactive or non-productive storage management tasks, and never allowed to address data archival and preservation proactively.
- o Data Access and Sharing - Most storage infrastructures and systems, though networked, are aligned to one or perhaps a few applications. This results in segregated silos of data, incompatible technology solutions, and disjointed policies rendering attempts to enforce consistent, company-wide communication and archival procedures ineffective.
- o Data Protection and Preservation - Although a wide variety of storage solutions and media types exist (tape, disk, optical) all have inherent liabilities that exacerbate IT's existing struggles with managing data growth, long-term preservation, and archival. To date, IT professionals have been forced, for the most part, to accept some level of compromise with their choice of archival media, as no one type provides the perfect set of characteristics.
- o Information Archival Solutions must be Purpose-built - Business and email applications, databases, and other productivity tools were developed to create, send, and receive electronic records, attachments, and other information. Out-of-the-box, even the leading applications don't provide the requisite features or functionality required to avoid or prevent threats against improper use of electronic data let alone enable users to conduct comprehensive yet quick search and retrievals of archived data, with original context, as needed.

- o Enterprise-wide Information Governance - The inability to enforce universally acceptable policies in a passive way, *aka* "the knowledge worker participation problem" – individual contributors are already burdened by the sheer volume of electronic information without the added expectation that they can and will follow passive archiving or governance policies during the course of normal business. A truly effective information archival solution will provide the ability to automate some time-consuming, non-value-add tasks, while still allowing users to access the data they need, when they need it, given appropriate security permissions..

Preventative, automated, minimally-invasive, and proactive policy-based management and governance of all corporate electronic records and other content is the holy grail of information lifecycle management (ILM) and represents a blueprint and best practice for IT and business leaders to follow as they take incremental steps towards enterprise-wide information governance. The key to achieving this level of complete control over all information assets is to place those same assets at the heart of any discussion that involves archiving. Organizations will find that as they address the tapestry of threats that face their business and IT operations, they must first adopt an "information-centric" view of their environment. In making the total collection of all corporate information that must be archived central to any discussion about compliance, governance, lifecycle, or simply storage management, organizations will begin the journey to transform current risks and liabilities into an true asset and competitive differentiator.

To Archive or Not to Archive?
Herein lies the rub – IT and business leaders know that they must improve their organization's ability to protect data through its lifecycle but cannot afford to actively archive "everything forever"; in fact if such a universal retention policy was financially possible, and technically practical, from a legal and governance perspective, retaining data that has outlived its usefulness or perhaps its mandated retention period, can be as dangerous as being unable to produce records when subpoenaed.

While tape and optical storage have been the choice for long-term data archival these traditional technologies may simply not support the performance, flexible online access, or speed-of-business recovery requirements today's enterprises demand. Though not a portable medium like tape or optical storage, new lower-cost disk-based solutions afford enterprises the ability to retain more data online in a readily accessible format. A disk-based archive strikes an attractive balance between offering a lower-cost alternative to expensive primary disk storage, while enabling a level of access performance that tape and optical storage cannot match, even with tape's historically lower acquisition cost. Furthermore innovative new types of disk-based solutions offer the equivalent of immutable or WORM (write-once, read-many) media, which directly satisfies certain regulatory requirements and further enables companies to maintain an authoritative record of the business.

Yet using disk storage to create an online archive for frequently accessed or perhaps compliant records is simply the first step. Without the policy-driven software to classify records, control which record classes are retained (and as importantly, not retained) the archive neither prevents intentional non-compliant activities from occurring nor provides assurance that your business is retaining the records that it should. Ironically, many organizations have adopted such a strategy of "keeping everything for ever" only to discover during an audit, regulatory investigation, or legal case that their archive contained incriminating or inappropriate communications which could have, and therefore should have, been permanently deleted. The conclusion we draw is that no single storage solution or media type can completely address all of the legal, operational, or financial requirements that modern enterprises and other businesses face. The real key for users, and subsequent opportunity for technology vendors, exists in integrating the most appropriate technologies into a flexible solution that meets tactical requirements while forming the strategic core of a purpose-built archive that is both actively accessible and intelligent.

Existing Archival Solutions and Processes Address Some But Not All Requirements
There are no shortage of mature and emerging technologies that promise to relieve some of IT's burden associated with archiving information, yet as mentioned previously, no single technology alone addresses all requirements. A short collection of some existing technologies and solutions which are being employed today for archival purposes include:

General-purpose and WORM Tape Storage
While tape storage remains an affordable and portable means to archive valuable corporate information, its serial nature renders it inappropriate for archiving data that is (or may be) frequently accessed. The ever-increasing pace of business is placing a greater premium on fast access to and sharing of data, two requirements that tape storage simply wasn't developed to provide. When an organization requires more tape storage capacity to archive its growing data repositories, it can simply purchase additional tape media providing a virtually unlimited capacity. However existing tape-based archival solutions do not provide a seamless level of access across both disk and tape storage, undermining the potential reuse of archived data and increasing the cost and complexity of management. While the leading tape formats are now available in the WORM variety where non-erasability and non-rewritability is required or desired to meet a regulatory guideline, it's still tape, and therefore suffers from the same inherent limitations.

WORM Optical Storage
Sharing many of the same positive characteristics as tape storage, optical remains a relatively affordable and portable means to archive corporate information. When an organization requires more optical storage capacity to archive its growing repositories of historical data, it can simply purchase additional optical media. That said, the storage and media management tools available for optical solutions don't match the ease-of-use capabilities of those available for tape storage. As a result, as the number of optical platters increases, so to does the cost and complexity of managing this resource and for organizations looking to build Petabyte-level active archives containing billions of data objects optical storage is not appropriate. Due to its inherent WORM attributes, optical storage has been *the* archival choice over the past 20 years for regulated organizations that needed to retain electronic records in an immutable format. However the explosion in data being created, and ultimately retained, coupled with the limited media and storage management tools that optical solutions offer, is putting pressure on optical vendors to remain relevant in today's environment.

General-purpose Disk Storage
The great advances in magnetic disk storage over the past 20 years have enabled the sum total of all historical information to be accessible online to anyone with an internet-connected computer. Yet while the cost to retain data on disk continues to decline, disk storage remains many times more costly than tape, forcing companies to make difficult and often inappropriate decisions as to which data sets are retained on disk and which are migrated off to tape. Furthermore, while general-purpose disk storage is appropriate for bulk retention of data, it doesn't offer granular, file-level management and protection or the WORM characteristics required to build an online or compliant active archive.

Hierarchical Storage Management (HSM)
While HSM is a relatively mature and well-understood technology, it is severely limited in terms of its ability to address the full range of information governance and lifecycle management requirements. HSM can address the need to automatically migrate older or less-frequently accessed data to slower-performing disk storage or perhaps offline to tape or optical media. That said, HSM solutions lack the ability to apply a value-based weighting to files and/or metadata, which would enable electronic records to be migrated not simply by age, but by relevance, purpose, risk, or other subjective criterion that the business creates. Furthermore, HSM solutions alone do nothing to address the myriad privacy, security, and compliance requirements facing organizations today.

Network-attached Storage (NAS)
Though offering the file-based management which is attractive to customers looking to build an active archive, the scalability of standard NAS storage is limited by the underlying file system and therefore can't grow to accommodate millions or billions of individual files or objects. Users will find that attempts to architect an active archive around a group of standard NAS systems are limited by the ability of the management software to accommodate massive growth, creating separate repositories of data, and management inefficiencies. While NAS solutions are more easily integrated with 3rd party applications by virtue of their NFS/CIFS interfaces, a general-purpose NAS system offers none of the native record and retention management features required for an intelligent active archive to deliver value beyond simply storing information.

Content-Addressable Storage (CAS) and CAS-like solutions

A relatively new concept in storage solutions is content-addressable storage (CAS), which, among other features, assigns a unique digital fingerprint to every unique file, sub-file, or object. The first generation of these disk-based solutions represent a compelling blend of online access, improved scalability and manageability over NAS, and compliance-enabling WORM functionality. Yet some of the currently-available CAS and CAS_like solutions do not provide a standard and/or open NFS/CIFS interface for easy integration with 3rd party and custom applications, while others offer no means to back the archival system up to tape/optical storage for offsite retention or disaster recovery purposes.

Standard Backup / Recovery Solutions and Processes

Many users still believe that their existing backup/recovery processes and solutions are also appropriate to provide the active and intelligent archival requirements their businesses now require. Standard backup/recovery solutions, however, were developed to make a copy of full or partial production data sets in the event of a system failure or data corruption, not to support the data access, data sharing, or storage management features of a true archive. Backup/recovery solutions tend not to be well integrated into the business-facing applications that create the data, and therefore provide none of the lifecycle management or workflow tools that many users are now looking for in their archive. While some IT professionals attempt to "kill two birds with one stone" – that is, by trying to satisfy the need to perform backups for data protection while also addressing the more business-centric requirements of an active archive – users must realize that these two processes, and the technology solutions associated with them, are unique.

Disaster Recovery / Business Continuance Solutions and Processes

If standard backup/recovery is inappropriately cast as a proxy for an intelligent active archival solution, expecting a disaster recovery (DR) or business continuity (BC) solution or DR process to address any of these same requirements is simply foolish. DR/BC should be viewed as a critical investment to make as insurance against a catastrophic systems failure or loss of primary data, destruction of a facility, or perhaps long-term interruption of utilitiy services.   (electricity, network connectivity, etc.). The technologies required to provide this level of protection, however, tend to be expensive and require a considerable investment *in anticipation* of a disruptive event, yet provide little if any value to the business during normal operations. Like backup/recovery, DR/BC provides the business with an alternative means to recover data if and when it (or the business itself) is destroyed or damaged, but also like backup/recovery, DR/BC should be viewed as a unique process and investment.

The (New) Rules of Archiving & Requirements for Building an Intelligent Active Archive
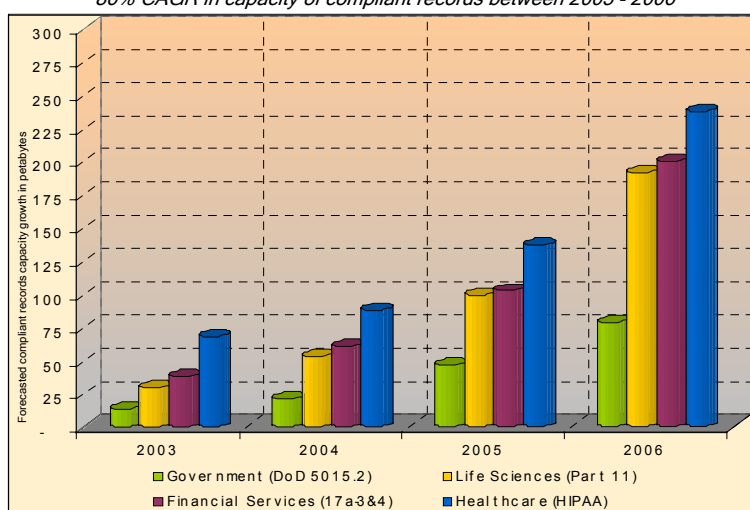
An Application-centric View of Intelligent Active Archiving

Since email is both a tool to communicate and also a transport mechanism for distributing, sharing, and collecting business-related information, it presents us with a perfect example to illustrate the benefits an intelligent active archiving solution offers. Specifically, how an intelligent active archive can provide the tactical benefit of helping organizations accommodate enormous growth in data, address specific regulatory compliance mandates (e.g., *retain all emails from certain individuals or of a certain type for 3 years on immutable media*) while also providing an effective means to capture, protect, and preserve records throughout each records' individual lifecycle (*retain 2 copies of email set "A" for 3 years on disk and email set "B" for 7 years on tape*). Ultimately, a truly integrated intelligent active archive will leverage the appropriate storage medium (*disk, tape, optical*) to satisfy both data protection and cost requirements while striking the right balance between providing robust, granular searchability of all archived data (*retrieve all emails and associated attachments sent to or from any Board of Director's member within the past 6 months*) regardless of the application that created the record or the actual physical device upon which the record is stored. In this way, the intelligent active archive enables the business to insulate itself against many of the risks discussed throughout this paper, while at the same time enabling valuable historical information to be accessed, shared, and reused for new or incremental opportunities as they occur.

Capacity of Compliant Records by Industry, 2003-2006

Worldwide scope; four major vertical markets

Life Sciences is the fastest growing industry in our study representing an 86% CAGR in capacity of compliant records between 2003 - 2006

Forecasted compliant records capacity growth in petabytes

Legend: Government (DoD 5015.2); Life Sciences (Part 11); Financial Services (17a-3&4); Healthcare (HIPAA)

Source - Enterprise Strategy Group, 2003

Ultimately, however, email is a single application representing a fraction of the total data a typical enterprise creates and retains; in looking at how an intelligent active archive can improve data protection, lifecycle management, and governance for this data type, business and IT leaders must not lose sight of the vision of true information lifecycle management (ILM) for all corporate data. In a way, active archiving and policy-based management of email is an application-centric form of ILM and provides a preview of the accountability, competency, and level of unified information governance that enterprises must build into their culture. The viral nature of electronic communications, and by extension, all digital information, requires information governance policies supported by an intelligent archival. Electronic records, must be identified, classified, and processed appropriately as close to the point of creation as possible without requiring every individual worker to manually ingest documents into a record or content management system or intuitively know what retention period to assign to each document they create

## Requirements for an Intelligent Active Archive

Storage-oriented Requirements

We are living in a world where access to the right information at the right time can mean the difference between winning and losing a contract or more importantly, saving a life. We are also witnessing an unprecedented number of innovative new technology solutions aimed at making the long-term retention of electronic information not only possible, but also cost-effective. To achieve an effective archival solution users must be empowered to select the appropriate technologies that meet their unique needs, yet some common attributes of such a solution can be defined. To summarize some of the requirements of a intelligent active archival solution relative to data storage and archival characteristics:

- o Accommodates massive scalability and unexpected growth – While a disk-based system with a massively scalable file system may offer the most attractive foundation on which to build an online archive, the complete intelligent solution will enable tape or perhaps optical storage to complement the disk for long-term retention and preservation when accessibility is no longer the highest priority. As it is likely that future compliance or regulatory mandates will require certain records to be retained indefinitely for decades or perhaps centuries, users must be able to choose when to migrate data from disk to tape, for instance, for both portability and cost efficiencies. That said, the intelligent active archive must be cognizant of data migration, using a combination of rich metadata and file system attributes to maintain an accurate account of where each record is currently stored to ensure timely access tomorrow or in 20 years.

Disk-based Solutions Address Tactical (Backup/Recovery) & Strategic (Active Archiving) Objectives

What are the key applications / business processes that you would likely support with a disk-based solution, rather than an enterprise-class tape library? (Percent of respondents, N = 94, multiple responses accepted)

Application	Percent
Backup/recovery	83%
Business continuity/disaster recovery	64%
Active (on-line) archive	44%
Video/multimedia	12%
Scientific research	11%
Medical imaging	7%
Other	4%
Don't know	3%

Source - Enterprise Strategy Group, 2003

- o Balance Technologies to Achieve Cost Efficiencies – The intelligent active archive should enable multiple storage mediums to be employed, with each being used to its most effective means. Disk storage is appropriate for online, rapid access, where high performance retrieval or collaboration is required, but this functionality comes at a price. Tape storage, while more cost-effective than disk storage, offers virtually limitless scalability and portability yet doesn't provide the high-speed random access characteristics many businesses and/or applications require.
- o Automate Manual IT Tasks – Provisioning new storage capacity remains one of the most time-consuming manual tasks that IT professionals must endure; a process that becomes increasingly cost- and time-prohibitive as data growth continues. An intelligent active archive automates certain manual tasks, such as provisioning new capacity, freeing IT resources up for value-add, incremental projects, yet still enabling the archive to maintain a high degree of autonomy during normal operations. The archive must also enable an order-of-magnitude increase in the amount of records, files, or data objects a single administrator can manage.
- o One Archive, Multiple Applications - By incorporating industry-standard interfaces for data management, such as NFS/CIFS, the intelligent active archive is more easily integrated with the
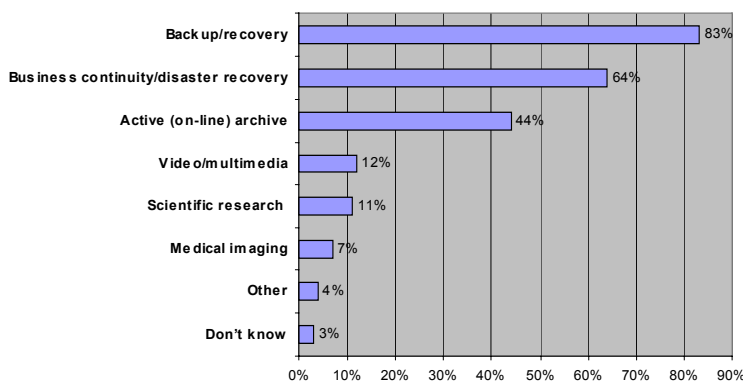
applications that create the records, files, and data objects that make up the history of the business. The intelligent archive then becomes less of a repository for old data, and more of a consolidated, indexed, and searchable collection of all the information the business has created, received, or shared throughout its life.
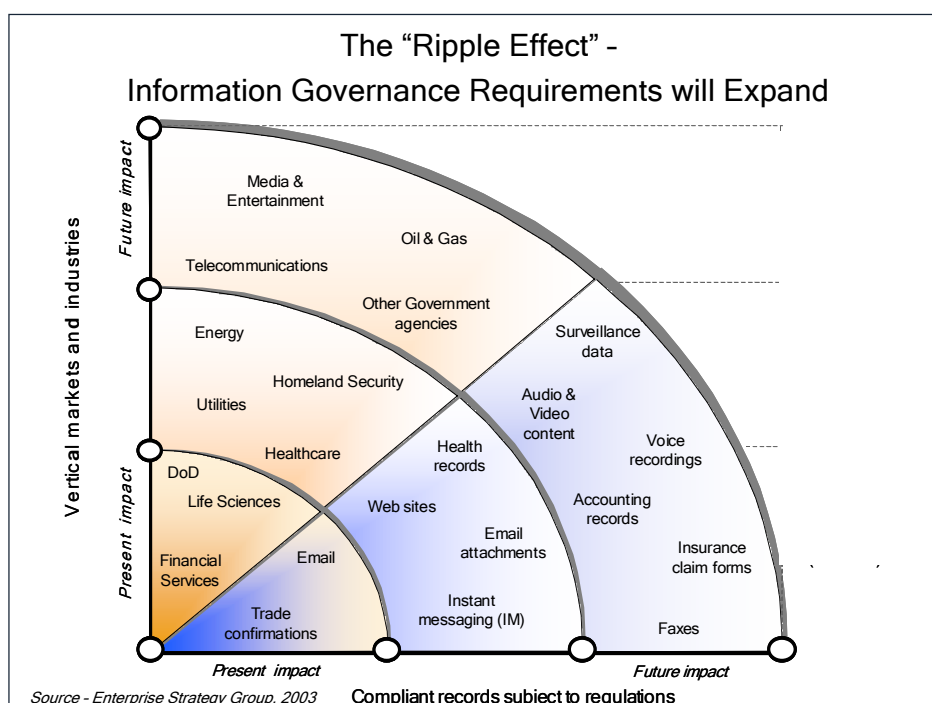
Information Governance-oriented Requirements

The old adage, "build it and they will come" is inappropriate when considering the purpose of building an intelligent active archival solution. Even if it were financially feasible for organizations to retain every unique digital record online within a disk-based archive, such a practice actually places the company in more risk – certain records should be permanently and securely disposed of after they've outlived their value to the organization or risk being discovered as part of a future, yet unrelated, electronic discovery event. To summarize the requirements of a intelligent active archival solution relative to governance, protection, and preservation characteristics:

o <u>Data protection and security</u> – Electronic records, files, and associated metadata should be protected with the appropriate levels of security, access controls, and encryption (where appropriate) to ensure compliance with internal corporate policy or regulatory guideline. Certain records should be protected in an immutable format on WORM media, while others can be retained on reusable media.

o <u>File or Data Object-level policy enforcement</u> – In much the same way a person's fingerprint is uniquely their own, every unique record, file, or data object should be assigned a unique means to identify it, independent of any file system description or physical location. Typically implemented via an algorithmic function within the archival system itself, this unique, digital "ID" provides users with the means to ensure the authenticity of data as it moves, is copied, shared, or perhaps, mistakenly deleted, throughout its lifecycle.

o <u>Immutable Creation, Preservation, and Disposition</u> – The intelligent active archive must be secure from both physical and electronic threats and is responsible for maintaining an undeniably accurate and verifiable account of all business records. In a business environment where corporate executives are held personally responsible to attest to the accuracy of their company's data, the "executive ignorance" defense no longer works.

Conclusion

 We are entering a new era of information archiving where the needs of the business to access, share, and maintain an accurate historical record of business transactions and electronic records meets or exceeds IT's standing mandate to simply provide "another copy" of data in the event of data loss or disaster. While we have only glimpsed the tip of the iceberg with regard to the impact that issues like regulatory compliance and corporate governance will have on the methods by which organizations manage information through its lifecycle, compliance is merely one of many drivers forcing IT and business leaders alike to reconsider the enormous scope of their archival requirements. Likewise, we still find ourselves in what amounts to the early morning hours of the Information Age – we've really just begun to understand information as a discreet asset, or the power of information sharing and collaborative development; none of which are able to be fully realized without a truly intelligent and active archive.   Information, being a dynamic and fluid element, must be managed, protected, and ultimately preserved and made available for future use by technological solutions that are intelligent, adaptive, and scalable.

The "Ripple Effect" –
Information Governance Requirements will Expand

Source - Enterprise Strategy Group, 2003

Both IT and business executives must view their role, at least in part, as stewards of their corporation's data. It's likely that more and more data will fit the profile of a "compliant record" and therefore must be treated appropriately, but again, simply addressing corporate governance misses the point.

If information truly is the new global currency, then the intelligent active archive represents the ultimate vault, meant to preserve, protect, and provide safekeeping for valuable information assets throughout their lifecycle. Corporate governance, regulatory compliance, and risk management all spring from the same source – that is treating information as the valuable and volatile asset that it is. It's never been more clear that legacy data protection and archival technologies and practices must evolve to accommodate the need to access valuable information now and for a long time to come. Taking an information-centric view of one's business is a sound starting point on the journey to transform information risk into a powerful competitive differentiator – enabling companies to truly unlock the full potential of their information assets.