

April 26, 2010

Strategic Assessment

Archiving and the Information Enterprise

Authors

Randy Dazo
Omri Duek

Published by

Dynamic Content Software
Strategies Consulting Service.

© 2010 InfoTrends, Inc.
www.infotrends.com

Abstract

From Netflix's movie recommendations to the optimization of traffic patterns in the public sector, organizations require more accurate and trusted information to innovate and make smarter decisions at a faster rate than ever before. Analysis tools and more powerful logic engines are at the core of this shift, adding a level of intelligence to organizations based on their historical data, decisions, and results. Underlying these tools, a well-established and comprehensive method to collect, decommission, and archive content that matters to the organization is necessary. This document details the basis for such an archive—one that securely and cost-effectively retains needed information, makes it readily accessible when needed, and safely decommissions what is not required for the organization.

For More Information

If you would like to order extra copies of this report, receive permission to use any part of the report, or be informed of upcoming market updates, reports, and related projects, please email us at info@infotrends.com.

Headquarters:
97 Libbey Industrial Parkway
Suite 300
Weymouth, MA 02189
United States
+1 781 616 2100
info@infotrends.com

Europe:
3rd Floor, Sceptre House
7-9 Castle Street
Luton, Bedfordshire
United Kingdom, LU1 3AJ
+44 1582 400120
euro.info@infotrends.com

Asia:
Hiroo Office Building
1-3-18 Hiroo, Shibuya-ku
Tokyo 150-0012
Japan
+81 3 5475 2663
info@infotrends.co.jp

Table of Contents

Introduction	3
From Archiving to Innovation.....	3
The Information Explosion.....	5
Content, Content Everywhere!	5
Multiple Formats, Multiple Channels, Multiple Problems	6
“Findability”	8
Case in Point: DuPont’s Costly “Keep Everything” Policy.....	8
A Holistic Archiving Strategy	9
Three Steps to Success	10
1. Think Big, Start Small	10
2. Be Honest About your Resources	10
<i>Appliances and Managed Services.....</i>	<i>11</i>
3. Find an Experienced Vendor Partner	11
InfoTrends’ Perspective	12

Introduction

Competitive organizations recognize that information drives better business, but these groups often do not realize the hidden value or potential opportunities that reliable information can expose. Now, more than ever, organizations are finding opportunities arising from the wealth of information currently available. They are leveraging information and business analytics for better and faster decisions, optimized processes, more predictable outcomes, and reduced risks. That said, organizations need to begin to think differently about archiving and the value it can provide. Rather than taking a siloed, ad-hoc approach to archiving, organizations should build a comprehensive archiving strategy that addresses immediate needs while positioning for future, long-term growth. A modular approach to archiving is the ideal solution, providing for unified collection and governance for one information source while providing the flexibility to add other content sources and meet future business requirements over time. Not only does this archiving strategy set the foundation for future innovation, but it also delivers immediate benefits in the form of hard cost savings and soft productivity benefits.

Unfortunately, most organizations have taken a reactive and ad-hoc approach to archiving. Performing simple “back-ups” to local or network-connected tape drives and other high-capacity storage, as active storage is filled, gives these organizations little insight into the underlying value of stored content (or the risks they might expose). This impromptu approach has bred an infrastructure of siloed solutions for e-mail content, file systems, structured content, network drives, and other media. As a result, these organizations have inadvertently created immense gaps in their information infrastructures, limiting the visibility of storage hard costs and access to (potentially) valuable historic data. With organization-wide content requirements on the rise, including high-cost and high-risk areas such as Records Management (RM) and eDiscovery, organizations can no longer afford to approach archiving in such a manner. Instead, a strategic approach is required. This assessment provides details on benefits and steps to success.

From Archiving to Innovation

Most organizations begin evaluating archiving tools or strategies due to a pressing issue or need. For many organizations and agencies today, the perennial growth of e-mail content can often serve as a starting point. Focusing on reducing storage costs and bringing order to chaos, IT managers have eyed tools that feature automated back-ups and advanced capabilities, such as de-duplication. These tools, which are focused on short term retention and recovery, do not provide the same value as archiving or expose the value of their stored content. Critical business information may be hidden in a constant stream of useless e-mails or petabytes of storage. It is here that a strategic approach to archiving, and granular content assessment capabilities, can provide greater value. By identifying valuable information for retention as well as decommissioning insignificant content, these tools provide the basis for more informed and proactive archiving. In other words, these tools add a level of intelligence to the archiving process.

A strategic approach to archiving implies more than just cataloging e-mails. This strategy must also include other sources of unstructured and structured information, such as databases, Customer Relationship Management (CRM) and Enterprise Resource Planning (ERP) applications, multimedia, as well as the rest of the vast landscape of business systems. In the same way that e-mail content must be assessed for business value, these systems must also be combed for useful information. Beyond the reduction of storage hard costs, technologies for unifying information siloes, extracting or enhancing their

business-critical data, and using that data in business processes have created incredible efficiency and innovation opportunities. Structured information from databases and unstructured content repositories can be married, analyzed, and acted upon faster and more accurately than ever before. Consider the following:

- In the healthcare sector, unstructured patient notes and structured clinical data are co-analyzed to improve the accuracy of diagnoses and contraindications.
- In financial services, unstructured communications and structured market data are used to automate and to inform highly-agile trading strategies.
- In the distribution and logistics industries, rapid access and understanding of information fuels “just in time” delivery and supply-chains.
- In the energy sector, real-time information collected from the U.S. “smart grid” will drive automation of energy production and routing during peak periods and for crisis situations, avoiding the brownouts and grid meltdowns of the past.
- In the travel industry, internal and external data sources are used to intelligently price offerings to maximize profits.
- In the advertising/marketing industry, data from Web, print, e-mail, and social media promotions are being used for multi-channel and highly-targeted campaign management and lead generation.

Even in consumer-facing industries (such as retail or media/entertainment), information is driving new innovations. For example, NetFlix’s “You might also enjoy...” feature uses historical customer data to identify the ideal DVD rental recommendations for its customers. Unknown to most, the algorithm also supposedly encourages greater distribution efficiency by skewing customer demand away from highly-demanded titles. A successful marketing and customer engagement tool, NetFlix’s complex use of operational information in delivering this feature also drives internal efficiency metrics.

An archiving strategy represents the first step towards an “information-driven” business philosophy by providing the foundation for information management, access, and process transformation. The information explosion has made archiving even more attractive for organizations today. Nevertheless, the increasing storage needs and complex information management requirements (such as Information Lifecycle Governance) are making the costs of an ad-hoc strategy impractical for on-going business. The need to have trusted information retention and consistent information disposal practices requires content assessment and analysis technologies at the core of an archiving strategy. By doing so, these practices enable organizations to understand the business value of the data that is being archived (keeping only what is necessary), determine how long keep the information relative to regulatory and/or business requirements, and decommission the data when it is no longer relevant. Although organizations may approach archiving by tackling their most pressing issues first (e.g., archiving e-mail to improve mail server performance and lower storage costs), the increasingly strategic role of information requires the evaluation of a product and platform flexible enough to meet future requirements mandated by the many information types continuously being conceived.

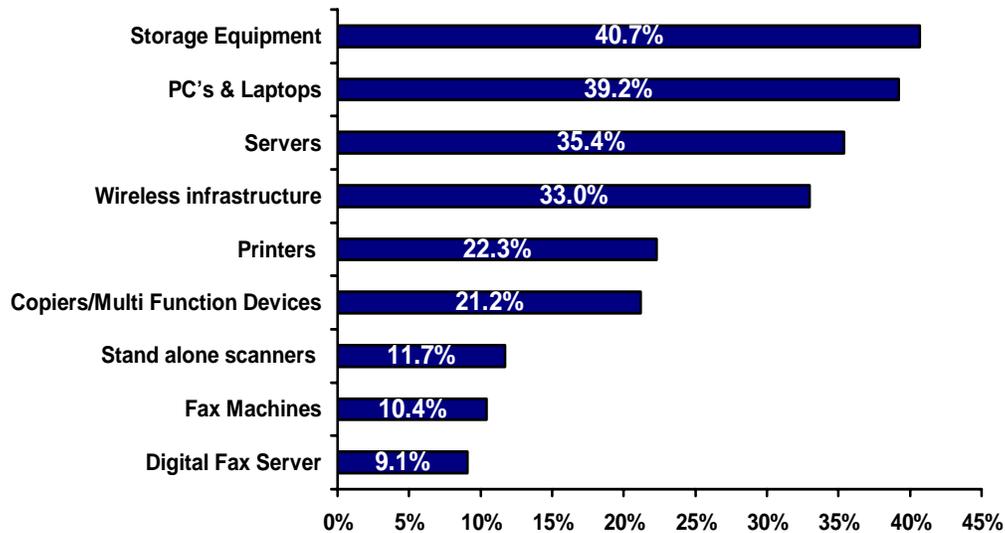
The Information Explosion

The recent market interest in archiving is not a shocking development for most observers. For years, organizations have approached archiving with a “hold everything” policy based on siloed back-up systems and few formal tools. Regrettably, these organizations must deal with very specific issues arising from the current information explosion, including:

- Exponential growth of content volumes and storage needs that eclipse the ongoing deflation in storage costs
- Increasingly complex information that can include structured and/or unstructured elements, multimedia, or collaboration suite feeds and data
- A growing need to differentiate the organization’s archiving policy with information that can be accessed anywhere, anytime, and on any device
- The inability to find and leverage information for future purposes
- High eDiscovery costs due to excessive material entering the eDiscovery process

Content, Content Everywhere!

The raw volume of enterprise information has become unsustainable. Although the amount of digital information in existence is subject to varying calculations, most research firms estimate it is currently on the scale of exabytes (1 billion GB) and will cross the zettabyte (1,000 exabytes) barrier by 2011. More important than the amount of information, however, is its cost to business. InfoTrends’ 2009 *Solutions and Services* primary research reveals that organizations across all vertical industries and organization sizes will be affected by the information explosion. With almost half of those surveyed responding that storage is the top area of increasing hardware investment.

Figure 1: Areas of Increasing Hardware Technology Investment (N=528)¹

Although storage hardware continues to deflate in price and alternative delivery methods (such as cloud storage) offer greater returns to scale, the raw growth in business content volumes exceeds these factors and forces organizations to increase their storage estimates as well as investments on an annual basis. These hard costs will weigh heavier on all businesses unless a formal strategy to determine the business value of stored content is developed and deployed. In the same 2009 survey, a surprisingly low 6.6% of organizations expect to *reduce* their data storage spend. As a consequence of “over-storing” irrelevant content, organizations will also inundate business users with information and, therefore, make it impossible to quickly locate trusted and actionable information for the task at hand.

Multiple Formats, Multiple Channels, Multiple Problems

In addition to the amount of content, the complexity of content is also increasing. Specifically, InfoTrends sees three factors that are driving richer requirements for content handling:

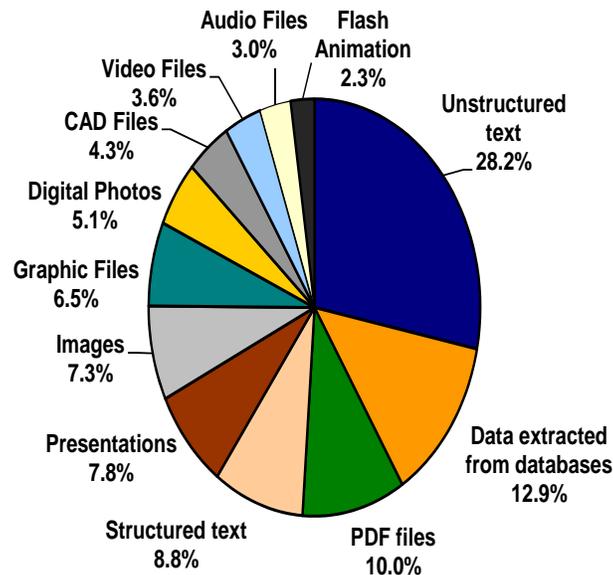
- Unstructured and structured information such as business communications and enterprise systems are increasingly becoming **multi-channel**, and information archiving strategies must account for these legacy and emerging channels. Unstructured information can include print, e-mail, Web, and mobile while structured information can come from ERP systems such as Oracle and SAP.
- Business content includes a growing amount of **multimedia**, such as images, video, audio, and Flash-like animation. This content requires the same storage and records diligence as other formats, although specific requirements and metadata for storing and accessing this media must also be met.
- No longer limited to e-mail, **collaboration suites** have transformed the way business processes are completed, leveraging social media techniques, real-time communications, and dynamic publishing tools. In all cases, these communications and content must be properly stored and linked with other relevant business documents and information.

Structured information and its management has become complex with organizations having multiple solutions satisfying certain business process and application needs. Organizations with multiple systems for ERP, CRM, WCM has become commonplace. Yet the organization and the federation of these

¹ Ibid.

systems is seldom addressed. InfoTrends' research of communications trends reveals that over 50% of *all* business content today has a multi-channel requirement. Moreover, the study revealed that over 25% of business content included multimedia components, excluding presentation content (7.8%) that often uses these embedded elements.²

Figure 2: Business Content, by Type (N=202)³



As visible in Figure 2, only about 20% of business information is associated with databases, data tables, or structured text. The vast majority of information is unstructured content, which makes finding and using this information significantly more difficult. Whereas typical data analysis and statistical tools can be used for assessing customer data tables, financial data, or operations results, the underlying unstructured content (e.g., customer e-mails, invoicing documents, and order receipts) is less accessible. Hidden within these unstructured documents, however, is often business-critical data:

- A customer's address or other contact information
- A purchase order, invoice, or vendor number
- Dates related to a project schedule
- Operational information, such as the cost of manufacturing resources
- Unstructured, textual explanations regarding a project delay or claim justification

As part of an intelligent archiving strategy, the value of this content must also be understood. Through technologies that can synthesize and analyze data, relevant information can then be added to unstructured content via assessment tools to make it usable, while insignificant information will be decommissioned.

² Franke, Maziarka, and Duek. *Multi-Channel Communications: The Content Publishing Workflow Challenge*. InfoTrends, Inc. March 2006.

³ Ibid.

“Findability”

More than ever, organizations are aware of the soft costs tied to information access and employee productivity. Siloed content repositories, separate systems, numerous user interfaces, and an inability to granularly search across the enterprise have stifled efficiency. Users are often forced to jump between multiple application “windows” as they search, cut, and paste information into a form that is relevant and useful for the process or decision at-hand. The information explosion has exacerbated this problem, and knowledge workers today are lucky to find important content in a timely manner.

Analyst firms studying the productivity of knowledge workers have quantified the costs of searching content. Although the methodology of these studies varies, the results indicate that the average knowledge worker spends 6-10 hours per week (approximately 1-2 hours per day) searching for information within the enterprise. Based on a burdened knowledge worker salary of \$60,000, this translates into organizational costs of \$9,000 to \$15,000 per year per employee. An organization with 1,000 knowledge workers that can save 20% of the time spent on search related matters could realize up to \$7.5M in worker-related productivity per annum. Of course, these figures become even more significant when taking into account higher-paid workers and those employees that conduct more searches than the average.⁴

Whether the need is to meet compliance and eDiscovery deadlines, to analyze historical data for present business decisions, or to empower analytics-driven business applications with trusted content, information access and analysis will define the next generation of competitive organizations. These data-driven businesses will have acknowledged the information explosion and deployed an enterprise-wide set of tools to identify valuable content as well as decommission insignificant content.

Case in Point: DuPont’s Costly “Keep Everything” Policy

Cheap storage hardware and a lack of strategy have led to a “keep everything” archiving policy at most organizations. Without an understanding of what content is valuable, or technologies for retaining and disposing of content in a consistent manner, organizations have defaulted to saving all their active and historical data. Duplicated documents and irrelevant or expired data reside side-by-side with business-critical content and compliance-related records. In addition to the additional storage costs such a policy imposes, the knowledge worker productivity loss—and, therefore, labor cost—can be enormous.

As an example, an internal case study conducted by DuPont in the early 2000s estimated that nine legal cases incorporated a total of over 75 million discovered pages. Estimating their costs at \$1 per page for processing and attorney review,⁵ DuPont realized that they had spent a substantial sum on these discovery processes. Perhaps more compelling, however, was another result of the internal study: DuPont found that approximately 50% of the pages reviewed were beyond their retention period, meaning that they could have decommissioned these records and avoided the discovery and litigation costs associated with these documents—a potential cost savings of almost \$12 million!⁶

⁴ IDC, Delphi Group, et al. Based on a 40-hour work week.

⁵ Based on costs for initial reviews of all documents (\$0.20/page) and attorney review of litigation-relevant documents (\$0.80/page).

⁶ J. Michalowicz. “Risk Reduction with Cost Reduction: A New Perspective on Cost Justifying ‘Best Practices’ ERM Program.” MER Conference. September, 2002.

The discovery of electronically-stored content is commonly used as a pre-trial negotiation tactic due to its cost and complexity. Knowing that the process can cost thousands (if not millions) of dollars between data restoration, data discovery, and legal review, defendants and plaintiffs used huge discovery requests to drive cases to early settlement. With the average Large American organization managing 556 cases concurrently, the cost of discovery has been a challenge for balance sheets—and DuPont was certainly no exception.⁷ On the bright side, the recently amended Federal Rules of Civil Procedure (FRCP) provides several safe harbor protections for organizations that do maintain their Electronically Stored Information (ESI) properly. In short, a compliant and consistent archiving and discovery strategy has become the difference between a successful and strong legal practice and windfall losses in legal fees, sanctions, and litigation (as DuPont’s own study has proved).

A Holistic Archiving Strategy

Unfortunately, because many organizations have taken the “Keep Everything” approach to archiving, not only do they take the risk of retaining expired data, but they also overemphasize the importance of hardware and hardware costs. A stronger strategy is a more holistic approach that consists of hardware, software, and professional services. It starts with a comprehensive assessment not only of an organization’s archiving needs, but also of its overall organizational goals. It also identifies the strategic role of information to the organization. Such a strategy would inherently include the following capabilities:

- Unified archiving for a range of structured and unstructured information systems, as well as a gamut of document, multimedia, and emerging content formats
- A comprehensive and constant assessment of the content that separates the unnecessary, duplicated and irrelevant from the trusted, information rich content that supports an organizations policies and governance objectives but also delivers the insight for better decision making
- A policy and decision making process through technology, that supports the governance and compliance initiatives that can serve the current and trusted content preventing human error and interaction
- A holistic program that includes continuous evaluation of the archiving ecosystem that help organizations reduce costs and at the same time provide better process, workflow and decision making improvements
- Mitigation of storage costs via content de-commissioning, de-duplication, and use-based storage medium selection (i.e., “Flash”-like storage, hard drives, and tape drives)
- Mitigation of compliance risks via information governance, as well as tools for litigation holds, privacy (i.e., redaction), and record integrity auditing
- Enhanced litigation readiness via proactive content discovery (i.e., search), assessment, and analysis
- Modular approach to start managing immediate needs, such as e-mail, file shares, and SharePoint archiving, but is also strategically flexible and scalable to include additional capabilities or content sources

⁷ Fulbright and Jaworski L.L.P. *Third Annual Litigation Trends Survey*. 2006.

A comprehensive archiving strategy and assessment seeks to reduce costs and increase productivity, primarily by normalizing the archiving and information access process across the enterprise, while reducing compliance and e-Discovery risks.

Three Steps to Success

To get started with a unified archiving strategy, InfoTrends offers the following steps to success for strategy development and partner selection.

1. *Think Big, Start Small*

Most organizations have inherited an “accidental architecture” of siloed and heterogeneous IT systems. Legacy systems, line-of-business purchases, and technology acquired via mergers and acquisitions (M&A) have challenged CIOs and IT managers when migration to a new system is not cost-effective or preferred. As a result, these systems are developed and supported in a siloed way— including the archiving of their content.

InfoTrends recommends conducting a thorough content strategy assessment and gap analysis prior to implementation of enterprise software. A successful deployment of an archiving program, or any major IT project for that matter, requires defining a targeted scope to get started. Whether the immediate need is to manage e-mails or bring order to chaotic file shares, addressing a specific and bound archiving issue (at first) allows for smoother introduction of strategic archiving across the enterprise. Nevertheless, when selecting archiving platforms and tools, organizations should consider the broader needs of the strategic initiative, including support for existing/emerging content types and systems, granular content assessment and analytics tools, as well as scalable support for growing content volumes. As the scope of the archiving strategy expands, support for mainframes, UNIX, Linux, and other platforms may become crucial. For maximum flexibility, consider vendors that adhere to open standards, such as Content Management Interoperability Services (CMIS), and have open and documented APIs to ensure that content can be integrated from multiple repositories. Doing so will help avoid creating islands of incompatible systems and disconnected information.

2. *Be Honest About your Resources*

Beyond content requirements, organizations must consider deployment requirements such as hardware, software, and professional services (e.g., planning, deployment, development, and maintenance). An organization trying to cut costs will often oversimplify these requirements during the planning phase. Unfortunately, such an approach leads to lower overall success rates and increased costs down the line as an organization must cope with “unexpected” costs and skill set requirements. That said, InfoTrends advises its clients to be frank and honest about their resources—especially in light of economic pressures.

For archiving projects specifically, InfoTrends recommends that organizations assess their capabilities across several areas:

- **Data Center and IT hardware:** capabilities (i.e., tiered storage, bandwidth) and capacities
- **Internal IT personnel:** skill sets and project resource availability
- **Project Management Organization (PMO):** IT-focused change management experience and project resource availability

Especially in a weak economy, where organizations may be running leaner than usual, InfoTrends expects that organizations will require third-party professional services to support deploying and maintaining their archiving strategies. Those organizations that lack archiving best practices experience should also investigate sourcing of planning expertise and support.

Appliances and Managed Services

Fortunately, more recent IT delivery models have streamlined options for organizations that need more than a set of install discs and a license key. Pre-configured appliances and hosted/managed services can help organizations cope with limited IT hardware or personnel resources—an increasingly common scenario given the information explosion and recessionary IT budgets. Appliances and managed services allow organizations to quickly deploy an archiving program (by bundling pre-configured hardware and software) as well as many of the front-end, labor-intensive configuration tasks without regard to IT hardware availability. For those organizations with limited IT personnel resources, a completely managed service may be the ideal choice as long-term maintenance and support are also bundled. Basic through advanced helpdesk, development and maintenance, as well as on-going platform monitoring services can all be outsourced to vendor experts as part of a managed contract—freeing costly IT staff from requirements that may be outside their expertise.

To minimize expense pressures, organizations are also looking at these options for their financial benefits. Transforming large, up-front capital expenditures to manageable operating expenditures, the Managed and Hosted pay-per-month models have become increasingly popular in the recession. IT initiatives such as archiving can benefit from these models, allowing organizations to quickly address content requirements with resource-sensitive financing and delivery. In many cases, these models make deployment possible where an organization could not otherwise conjure up the necessary up-front costs or internal resources. As an organization evolves and its information requirements change, it may choose to bring capabilities in-house or migrate them from appliances to formal data center servers. Choosing a solution that uses the same software platform for hosted, on-premise and appliance offerings will provide consistent functionality, flexibility and information access across delivery options. This is paramount for organizations that may want to move seamlessly between hosted and on-premise delivery options.

3. Find an Experienced Vendor Partner

At the vendor evaluation and selection stage of archiving deployment, look for an organization that has the experience to provide a comprehensive archival strategy, as well as the ability to help make better business decisions and find opportunities through innovation. InfoTrends advises that organizations seek a long-term archiving *partner* and not simply a *vendor*. We have identified several areas where the relationship with the prior is differentiated. An experienced vendor:

- conducts a comprehensive assessment of the current archival goals and infrastructure, but also looks to understand how a holistic archival strategy can help organizations with reducing costs or creating new opportunities.
- provides an open architecture that can unify the content infrastructure including support for legacy systems, such as mainframes and emerging technologies for rich media and collaboration content. Support for cloud deployment should also be considered, where optimal.

- offers a complete archiving portfolio, including integrated tools for compliance, content assessment and analysis, and e-Discovery. These tools integrate across the IT stack to help govern information over time based on business value.
- has a significant amount of vertical- and business-specific expertise, evidenced by prior successes and thought leadership with other customers.
- has a customer-centric approach to metric development and tracking, mapping their customers' internal goals to their own in all pre-sales, sales, planning, deployment, and support activities.
- successfully manages “global” operations to support its customers across regions, and “local” operations to understand regional business requirements and cultural differences.
- takes a consulting-led approach in contrast to a sales-led approach. The approach often incorporates a formal improvement methodology, such as Six Sigma or TQM.
- offers flexible delivery models (e.g., on-premise, appliances, or managed services) and financing options that meet its customer's IT and financial requirements.
- is viable for long-term archival and innovation strategies, a heavy consideration in markets where acquisitions and consolidation are occurring at a fast pace.

In contrast, *traditional* vendors are more interested in “moving boxes,” an approach that could be costly and disastrous for a strategic IT initiative like archiving.

InfoTrends' Perspective

Organizations today already differentiate with their information management strategies, but emerging technologies for assessing and analyzing structured and unstructured information are on the verge of revolutionizing this era of information usage. By extracting critical concepts and metadata, these technologies can catalyze smarter and more-automated workflows as well as faster, more comprehensive accessing of information.

Although the benefits of these emerging information technologies are applicable across virtually all content processes, archiving represents “low-hanging fruit” for organizations to address. Hard costs associated with storage and records managements (RM) can be quickly contained, soft costs associated with information access and productivity can be reduced, and risk associated with litigation readiness and compliance can be mitigated. Moving forward, this unified collection approach will also build the foundation for other data-driven innovations. That said, matching an organization's business, IT, and financial needs to a set of hardware, software, and services is the key to recognizing benefits quickly and successfully.

Therefore, a strategic, enterprise-wide approach with practical starting points is required, and InfoTrends recommends that organizations follow the following steps to success:

- Start with a pressing and targeted archiving need, such as e-mail management, but do not ignore broader and long term strategic requirements, which may include archiving structured and unstructured information.
- Be honest about internal IT and change management resources, including hardware and professional time.
- Choose a vendor that satisfies content as well as organization requirements for long-term viability and delivery models, including on-premise, appliances, or managed services that fit the organization's needs.

This material is prepared specifically for clients of InfoTrends, Inc. The opinions expressed represent our interpretation and analysis of information generally available to the public or released by responsible individuals in the subject companies. We believe that the sources of information on which our material is based are reliable and we have applied our best professional judgment to the data obtained.