

## WHITE PAPER

---

# Enabling SharePoint Operational Efficiency and Information Governance

Sponsored by: EMC

---

Laura DuBois

Marshall Amaldas

January 2012

## EXECUTIVE SUMMARY

This white paper is dedicated to providing firms with guidance on addressing the operational and information governance challenges associated with managing SharePoint content sprawl. This white paper will aid the reader by highlighting the immediate operational value and information governance benefits associated with taking control of both active and inactive SharePoint content.

This white paper highlights the benefits the EMC SourceOne family of information governance solutions provides in more effectively managing SharePoint growth by improving performance, reducing infrastructure costs, and managing SharePoint content according to legal and regulatory mandates. Using the SourceOne family of applications allows firms with record retention or litigation readiness objectives to effectively apply retention rules for content and/or sites as well as proactively respond to eDiscovery or investigative preservation and collection requests for SharePoint content. Firms can also archive inactive SharePoint content as a cost-effective, compliant, and consistent way to deal with the content sprawl that the use of SharePoint introduces.

## SITUATION OVERVIEW

According to *Archiving in the Context of Information Management and the Emerging Use Case for Archive Information Reuse* (IDC special study #225384, November 2010), approximately 72% of the 508 firms surveyed in 2010 were already using SharePoint and only 20% had not yet deployed the application. SharePoint adoption is also increasing as more organizations are making net-new investments in SharePoint after the release of SharePoint 2010. While the number of SharePoint installations continues to increase, the volume of data stored in these repositories is also increasing.

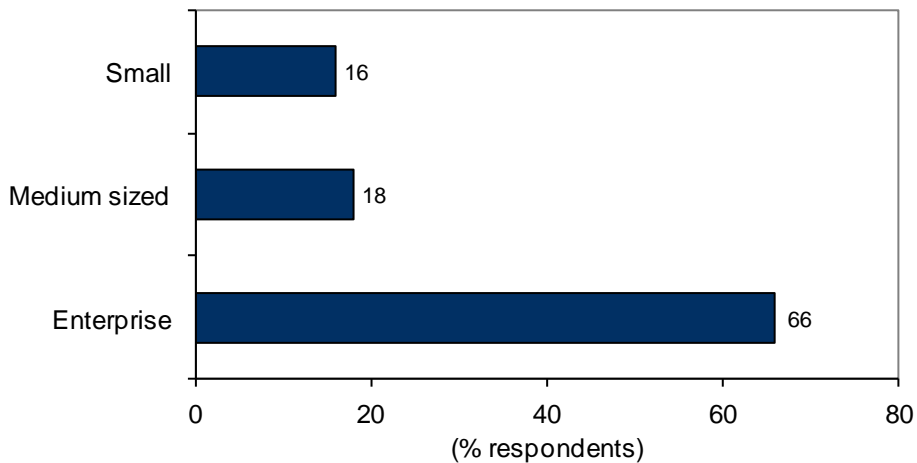
A 2011 IDC *Quick Poll Survey* on archiving revealed that a sizable fraction of organizations that use SharePoint and other ECM solutions retain their SharePoint data well over the duration of time that the data is used for meeting current collaboration needs or for supporting business workflows. According to the survey, 37% of organizations retain information in these repositories for two to seven years, and another 34% indicated retaining this information for up to seven years or more. While some organizations meet their retention needs by archiving information in a separate infrastructure, many still archive the data in-place within the primary SharePoint infrastructure.

Approximately three-fourths of all SharePoint data today tends to be large BLOB (binary large object) content. Attempting to manage and retain this BLOB content (document, images, rich media, etc.) in-place within SharePoint severely burdens the SharePoint environment. This approach not only is detrimental to the performance of SharePoint because of the impact on SQL servers but also steeply increases SharePoint management challenges for IT. Such burdening and slowing down of SQL servers' performance in the SharePoint production environment is detrimental to workflows and end-user experience.

Supporting this premise is IDC's 2011 *Trends in File-Based Storage Survey*, which revealed that almost one-quarter (23%) of the 261 survey respondents identified controlling proliferating SharePoint instances as a top file-based storage management issue. Among this subset of respondents, a majority (66%) were from enterprise size firms (see Figure 1). The mere size and number of business units and departments make it particularly challenging to devise and align data management and retention policies across SharePoint sites. This leads to indecision and suboptimal information management of SharePoint environments, which ultimately leads to larger amounts of data retained over long periods of time.

**FIGURE 1**

Firms Considering SharePoint Proliferation as a Top File-Based Storage Management Challenge by Company Size



Note: Data is based on a subset of respondents who chose SharePoint as a top management challenge.

Source: IDC's *Trends in File-Based Storage Survey*, 2011

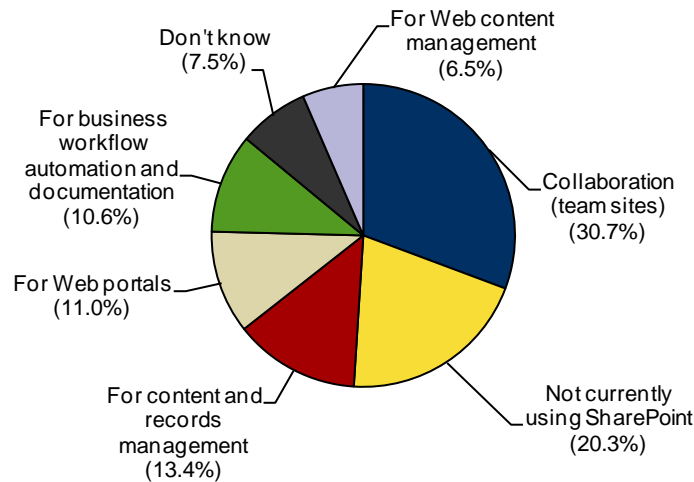
The predominant use cases contributing to the growth in SharePoint are collaboration, content, and records management (see Figure 2). The initial adoption of SharePoint is frequently accelerated by accessibility, usability, and affordability reasons. As a result, SharePoint adoption often starts as a business-led initiative at a departmental level within a large organization and often outside the purview of a

corporate IT function. Thus initial deployments frequently lack proper architectural planning, IT processes, or corporate governance oversight.

However, as more users, teams, and departments begin to incorporate the use of SharePoint into their work processes, for example using the tool to communicate and share information, scalability and performance challenges with initial implementations quickly surface. To address user experience requirements, system response times, or noncompliant information management methods, an architectural team is brought in to analyze and rectify the problems.

**FIGURE 2**

### Use Cases for SharePoint Deployments



n = 508

Source: IDC, 2010

### SharePoint Challenges

SharePoint growth in content, number of sites, and users brings several challenging realities to an organization. These operational and information governance related challenges include:

- ☒ **Performance and scalability:** For optimal SharePoint performance, Microsoft sets size limitations on SharePoint content databases. Subject to the performance characteristics of the underlying disk system and meeting certain standards pertaining to storage availability and disaster recovery capabilities, the recommended size limitation can vary.

For a general usage scenario, Microsoft recommends a threshold of 200GB for content databases in SharePoint 2010. As soon as the SharePoint SQL

database reaches this threshold, the SharePoint application response time starts to degrade. Performance not only degrades on response times but also leaves less processing power for indexing processes, backups, or returning search results.

- ☒ **Backup and recovery:** In addition to setting limits on SharePoint content databases, Microsoft also sets thresholds for SharePoint site collection in order to have a granular site collection–level backup and recovery capability. Microsoft recommends limiting the size of site collection to 100GB as operations such as site collection backup and restores weigh heavily on Microsoft SQL server, which can ultimately lead these operations to fail. When the size of the site collections cannot be contained to the recommended threshold, backup and recovery operations need to be performed for the entire content database. This not only affects the speed of the backup and recovery process but also hampers the precision of recovery when files are accidentally deleted and need to be retrieved.

Last, when the size of the database exceeds the 200GB content database limit, Microsoft's native SharePoint 2010 server backup is not guaranteed to meet backup and restore requirements. Beyond this limit, organizations may have to consider purchasing additional backup software licenses.

- ☒ **Infrastructure and operational costs:** To scale large SharePoint implementations, the SharePoint architectural components (SharePoint server, IIS server, SQL server, and media server) are deployed on separate physical machines, which can increase cost and management overhead. It can also mean increased complexity and time in performing administration tasks.

Second, many firms use SharePoint for project-, product-, or customer-specific collaboration. This collaboration may have a definite start and end period where content in the SharePoint site is no longer being used, accessed, or referenced. Maintaining inactive site content within a SharePoint instance is not operationally effective. SharePoint does offer the ability to delete inactive sites, but most firms do not want to destroy their intellectual property; instead, they prefer to retain it for possible reuse, reference, or request.

- ☒ **Preservation and search efficiency:** Finding content within SharePoint is possible using its native tools. However, eDiscovery requests, internal investigations, and audits are commonly done by a topic at hand or custodian, requiring a horizontal approach across systems. Separate, redundant processes for preservation, discovery, and collection of content in SharePoint are costly and may also compromise consistency and defensibility. Last, even when using the native SharePoint search functionality, an unpruned SharePoint database requires more information to be parsed through when performing searches, which ultimately impacts search performance.

- ☒ **Corporate visibility and retention:** Typically, initial SharePoint implementations are often outside of corporate IT visibility or control, with the growth of SharePoint sites and usage in either project or team scenarios. As a result, IT organizations and governance professionals alike are frequently not cognizant of

the corporate content that is stored within SharePoint data stores. Information within SharePoint commonly represents institutional knowledge for a firm and may be subject to retention rules but often this data is not managed according to the overall records management objectives. Native retention capabilities within SharePoint often are not integrated with broader records retention strategies or information governance policies. Compliance, records, or information governance professionals therefore need to apply user-, content-, or site-specific retention rules in a siloed fashion.

## **THE OPERATIONAL AND INFORMATION GOVERNANCE BENEFITS IN GAINING CONTROL OF SHAREPOINT GROWTH**

As previously discussed, the proliferation of SharePoint raises both operational- and information governance–related challenges. IDC research shows that enterprises are increasing the size and scope of their collaboration environments, with SharePoint emerging as a major driver of information growth.

*Microsoft SharePoint Server Ecosystem and Customer Usage Trends* (IDC special study #207413, July 2007) surveyed 300 U.S. organizations about their current and planned usage of SharePoint within their organization. SharePoint deployments play a major role in extending information access to a greater number of employees and opening access to a wider set of enterprise data assets. For large firms, SharePoint sites easily encompass thousands of individual contributors and hundreds of different workspaces, often deployed in a nonuniform fashion and initially outside of centralized IT architecture and planning.

As SharePoint expansion occurs, firms must develop a centralized information management strategy and standard deployment architecture to ensure data integrity, long-term scalability, and cost management. In gaining control over SharePoint infrastructure and content, firms not only achieve operational efficiencies but also enable broader information governance objectives.

---

### **Operational Benefits in Controlling SharePoint Content and Site Growth**

Information governance stakeholders such as legal, compliance, risk management, human resources, audit, and IT governance officers and professionals will gain profound, actionable information governance advantages. These include:

- ☒ **Cost-effective control of inactive sites:** Depending on the usage, SharePoint sites can become inactive over time while taking up unnecessary server infrastructure. This may be because of completed projects, making the site obsolete, or sites that are no longer needed. SharePoint allows for sites to be deleted, although the content may need to be retained for institutional knowledge or governance reasons. Archiving of SharePoint sites enables site retirement, freeing up SharePoint and server resources. Archiving also enables tiered storage to be leveraged, further reducing storage costs.

- ☒ **Retention management:** Firms can enable policy-driven and consistent retention and disposition of rules to SharePoint content based on different object properties. This ensures content is kept only for as long as prescribed by regulatory mandates. Retention can be applied within SharePoint itself, but using an independent archive enables holistic retention across different content types. A centralized archive also streamlines administration by enabling policy consolidation and consolidated searching across all archived content.
- ☒ **Search efficiency:** SharePoint does enable search with its Enterprise Search capability. However, in eDiscovery and investigative processes, SharePoint is commonly a secondary target to email and files. Legal teams want and need to execute searches horizontally across systems rather than system by system. Thus leveraging search capabilities across different types of archived content using the same processes, keywords, and people increases workflow consistency and defensibility while mitigating the risk of human error.
- ☒ **Visibility gained:** Gaining insight into SharePoint content benefits information governance stakeholders that are required by the firm to ensure compliance with regulatory retention, information security, data privacy, and legal discovery objectives.

In addition to meeting retention management, eDiscovery and investigative readiness objectives, firms achieve operational objectives by leveraging an information governance approach to SharePoint.

## **EMC SOURCEONE FOR OPTIMIZING PERFORMANCE, COST, AND INFORMATION GOVERNANCE PROCESSES**

EMC SourceOne is a family of information governance products that make information governance actionable. It includes application optimization, archiving, and eDiscovery solutions that help address customers' most pressing information management needs. These products are modular, giving customers of all sizes "focused solutions" to start with most of their challenges — to deliver immediate business benefit — and then expand over time. EMC SourceOne allows proactive, consistent, and repeatable management of retention and disposition policies and, as appropriate, long-term preservation based on the value of the content. The SourceOne family of products boosts application performance, enhances operational efficiency, enforces retention and disposition, and leverages tiered storage to streamline operations and reduce costs.

According to EMC, information governance is designed to lower costs, reduce risk, and ensure compliance with legal, regulatory standards, and/or corporate governance. Information governance includes policies and technology to understand what information is at what point in its life cycle and to apply the appropriate policies, including retention, disposition and, as appropriate, long-term preservation. It includes visibility into the information within the organization, allowing the organization to understand what information it has as well as where that information is stored and to take action on the information.

Information governance encompasses the people, practices, and technology to proactively manage and take control of information:

- ☒ **What** information is stored (classification, information visibility)
- ☒ **Where** information is stored (tiered storage, in-place management, and/or secure legal hold)
- ☒ **Who** has access (role-based access for eDiscovery; broad user search for general office productivity)
- ☒ **How** long information is retained and preserved (includes retention classification, records management, WORM storage)

The EMC SourceOne family of products and solutions currently features:

- ☒ **EMC SourceOne Email Management** provides all core email archiving capabilities for Microsoft Exchange, IBM Lotus Notes/Domino, and instant messaging. This new generation of email archiving is built upon a next-generation platform that supports modular expansion of the application to archive a broad range of content types including SharePoint and file systems.
- ☒ **EMC SourceOne Email Supervisor** complements the SourceOne platform as an optional add-on. This application provides for cost-effective monitoring policies against emails to allow customers to validate adherence to corporate email usage policies. Flexible sample rules and lexicons allow customers to monitor as broadly or narrowly as they wish.
- ☒ **EMC SourceOne Discovery Manager**, also a SourceOne platform additional add-on, discovers, manages, and applies secure legal hold to content in the EMC SourceOne Email Management archive. With Discovery Manager, organizations can quickly find, safely hold, efficiently cull, and defensibly produce archived email in response to a legal or regulatory notice or a corporate policy complaint. Built around a legal matter or case metaphor, Discovery Manager supports secure authorized investigator access, defensible collection results, and chain of custody.
- ☒ **EMC SourceOne eDiscovery–Kazeon** provides defensible in-place assessment of content and collection (if required) from desktops, file servers, archives, email servers, and other repositories that exist "in the wild" and outside of a formal archive. This solution allows for the discovery, indexing, analysis, and collection of content stored in a variety of repositories and satisfies the need to quickly respond to eDiscovery, eDisclosure, audits, and investigative requests.
- ☒ **EMC SourceOne for Microsoft SharePoint** supports and enhances an organization's use of Microsoft SharePoint by optimizing the application for enhanced performance, reducing storage costs, and extending control and governance to SharePoint content without impacting the end user's experience.

## **EMC SourceOne for Microsoft SharePoint**

EMC SourceOne for Microsoft SharePoint enables firms to improve operational efficiencies, system scalability, and performance while reducing the burden and costs incurred in the IT department. By providing compliance, operational control, and discovery capabilities, EMC SourceOne for Microsoft SharePoint helps organizations meet regulatory requirements and reduce corporate risk.

### ***A Building Block Approach to Managing SharePoint Content***

EMC SourceOne uses a building block approach to help firms manage the growth of active SharePoint content. By using content externalization, firms can reduce the burden on SharePoint and can immediately improve SharePoint's performance, which ultimately leads to better operational efficiency and reduced costs. Organizations can also choose to archive inactive content to a central archive, by which IT managers can extend the control and governance to SharePoint content to mitigate legal risks

- ☒ **Getting started with SharePoint:** SourceOne helps a firm achieve *operational efficiency* by reducing cost, improving data protection, and increasing scalability and performance.
- ☒ **Scaling existing SharePoint deployments:** Firms can gain *visibility into and control over current SharePoint* content and sites. SourceOne allows a firm to effectively control current site and content growth and manage inactive sites, according to policy and without disrupting users.
- ☒ **Mitigating risk with SharePoint:** Firms can ensure consistent and defensible *information governance* for SharePoint content by ensuring compliance and eDiscovery readiness. SourceOne supports the archiving and retention of SharePoint content into a SourceOne archive that is searchable should the need arise to search and produce content in response to an investigation or litigation.

### ***Benefits of Using SourceOne for Microsoft SharePoint***

The SourceOne for SharePoint application has a breadth of features that offer firms the following operational and information governance related benefits:

- ☒ **SharePoint performance and scalability gains:** SourceOne for SharePoint supports policies for externalizing content from SharePoint production systems directly to storage platforms. Externalization enables SourceOne to move the content from the SQL database to lower-cost tiers of storage without disruption to content access from within SharePoint. This improves performance in the production SharePoint system and extends the scalability of the SQL servers.
- ☒ **Storage optimization through tiering:** This feature offers the ability to leverage a tiered storage infrastructure for managing SharePoint content, which can greatly reduce storage costs. IT managers can follow a multipronged approach to managing SharePoint storage cost by moving inactive SharePoint content from expensive, highly availability storage devices to low-cost storage while externalizing the active SharePoint content directly to higher tiered storage.

Following this methodology not only reduces storage cost but also reduces the burden on SQL servers.

- ☒ **Proactive SharePoint/storage health monitoring:** EMC also provides additional options for SharePoint such as a health monitor to ensure that the current IT resources are optimally utilized while eliminating the risk of errors during the externalization process. The health monitoring feature allows IT managers to set thresholds for the size of the SharePoint content database and storage. Monitoring the content database helps ensure that the SharePoint environment is performing optimally, while monitoring storage helps avoid the risk of running out of storage capacity.
  
- ☒ **Consistent, enforced records retention:** SourceOne for Microsoft SharePoint supports two types of policies — archive and archive and delete — for content in the SharePoint production environment. Both archive and archive and delete actions will apply a preconfigured retention policy to the content. An archive action moves the content from SharePoint without impacting user access. Conversely, archive and delete moves the content to SharePoint and deletes the content from the SharePoint site completely. All SharePoint content including objects, document libraries, form libraries, picture libraries, entire SharePoint sites, discussions, lists, contacts, Wikis, and tasks can be archived using these policies. SourceOne for Microsoft SharePoint imposes no restrictions on the type of content that can be archived. Searching for SharePoint content is done from SharePoint for end users, but administrators can perform a global search of all content via the SourceOne Web search interface.
  
- ☒ **Transparent process for users and IT help desk:** Regardless of whether a user is accessing archived data from the SourceOne archive or active data that is externalized via RBS, the back-end processes are transparent to the end user. SourceOne does not rely on or maintain stubs or shortcuts to SharePoint content. SourceOne for Microsoft SharePoint leverages Microsoft-recommended APIs to offer a transparent and seamless archive experience to the user. This eliminates the need for IT to support help desk calls from users trying to manage their shortcuts. The common problems of shortcutting, such as renaming, moving, or deleting shortcut or stub files, are eliminated. Additionally, indexing overhead can be minimized by eliminating the need to index or re-index shortcut files.
  
- ☒ **Seamless SharePoint user interface and information access:** SourceOne for SharePoint provides a seamless end-user experience for content it archives and externalizes by allowing users to access content from the SharePoint native UI. Users do not need to go to a separate portal or interface to access content. Additionally, SourceOne preserves the attributes associated with structures in SharePoint. For example, SourceOne archives, stores, and returns to the SharePoint user the content type in context (e.g., the difference between a contact and a task). The result is the archiving process is seamless from the user's perspective. Users do not need to be cognizant that content they access lives outside of SharePoint.

- ☒ **Flexible, extensible archive and retention policies:** SourceOne for SharePoint can archive content based on properties such as creation date, last modified date, and file extension as well as custom, user-defined metadata or the query results from a Collaborative Application Markup Language (CAML) search query. The benefit for the user is the flexibility in customizing the archive and retention policies based on either a typical time-based retention rule or a metadata field specific to a particular firm's business industry or environment. Last, administrators can execute a CAML search with the results from that search either on demand or programmatically serving as the source for a specific retention policy. This gives the firm a large degree of flexibility in applying the policy best suited for its environment and information governance needs.
  
- ☒ **Breadth of versions supported:** SourceOne for SharePoint supports SharePoint 2010, SharePoint Foundation 2010, MOSS 2007, and WSS 3.0 support.

EMC SourceOne for Microsoft SharePoint has knowledge of SharePoint object metadata and content. This combination is imperative to understanding not only the intrinsic value of the data as an intellectual property but also the risk associated with loss or misuse of such information. As a growing portion of enterprise content is stored or organized within SharePoint, this awareness of SharePoint data sets is critical. Such awareness needs to be accompanied with capabilities to enforce intelligent policies for migration, classification, search, tiering, preservation, and retention based on content-triggered rules.

## CHALLENGES

As the economy continues to be sluggish and unpredictable, organizations across the board are working toward maximizing the utilization of their existing IT assets in order to deal with constrained budgets. IDC expects IT spending for the next 12 month to be focused around technologies that enable higher efficiency gains and IT asset consolidations that help minimize underutilization of compute and storage resources. In most organizations, optimization of SharePoint through externalization and archiving presents a good opportunity to economize on IT infrastructure spending and is often seen as a low-hanging fruit for organizations serious about boosting IT efficiency.

Second, one of the questions firms face as they deploy or scale SharePoint is how does SharePoint fit, if at all, with a firm's existing investment in content repositories and/or records management systems? In practical reality, SharePoint introduces yet another repository to manage and ensure information governance for. SharePoint is yet another system to apply records retention rules to and conduct eDiscovery preservation and collection against. Managing multiple repositories according to operational and information governance objectives can be problematic as a firm grows in size and may ultimately become a source for process inefficiencies.

The challenge firms and the industry at large face is providing adequate levels of integration across these disparate content repositories, which increasingly include SharePoint workspaces. However, in the long term, this integration is needed to ensure a centralized, consistent, enforceable, and repeatable process and policy for information indexing, archive, retention, search, disposition, and destruction and to make information governance achievable.

The EMC SourceOne strategy is to work with (and not override) retention policies placed on active SharePoint content while managing only inactive SharePoint content or SharePoint content that is explicitly moved into an archive. However, any SharePoint content that is externalized but remains active content is still managed under policies set within the SharePoint application. EMC's approach ensures policies are executed as expected while also offering operational benefits.

## **CONCLUSION**

As SharePoint continues to be widely adopted, it has become a focal point for both information access and aggregation as well as a project collaboration point where documents are created, revised, and shared. The volume and importance of the information contained on SharePoint sites also heighten concerns about operational efficiency, information governance, and data retention practices.

Business workers want assurance that their files will be quickly accessible, searchable, and recoverable at will. IT managers want to cost effectively optimize the SharePoint environment to reduce storage costs and meet required service-level agreements. Information governance stakeholders such as legal, compliance, risk, and records officers must ensure that information contained within SharePoint sites is retained according to regulatory mandates, properly preserved pursuant to a legal hold order, or effectively searched as business, legal discovery, or audit demands dictate.

The SourceOne family of information governance products enables a firm to gain control over its SharePoint content. SourceOne for Microsoft SharePoint allows for SharePoint growth while improving performance and reducing infrastructure costs. It also ensures a firm can meet legal and regulatory mandates associated with SharePoint information. By ensuring a central repository of archived information, administrators can quickly search for information that is requested in response to an investigation or litigation. EMC SourceOne eDiscovery–Kazeon enables SharePoint content collection from SharePoint repositories "in the wild" for the purposes of supporting eDiscovery and broader event-driven investigation workflows. The EMC SourceOne family of solutions allows a firm to achieve its operational and information governance objectives for SharePoint content.

---

## **Copyright Notice**

External Publication of IDC Information and Data — Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Copyright 2012 IDC. Reproduction without written permission is completely forbidden.