



INFORMATION INTELLIGENCE FOR MICROSOFT SHAREPOINT



Mercy Ponnupandy
Senior Practice Consultant
EMC Corporation

Table of Contents

Introduction 3

SharePoint Capabilities 3

SharePoint Implementation Challenges 4

 ECM Challenges 4

 Operational Challenges 4

 Compliance Challenges..... 5

Intelligent Enterprise Capture 5

Intelligent Information Access 7

 Content Management Interoperability Services (CMIS) 9

 My Documentum for SharePoint..... 10

 EMC Documentum Repository Services for Microsoft SharePoint 12

Intelligent Information Governance..... 14

Conclusion 16

 Improve Performance and Reduce Costs 17

 Scalable, Flexible, and Extensible 17

 Ensure Compliance 17

Appendix..... 19

Disclaimer: The views, processes, or methodologies published in this article are those of the author. They do not necessarily reflect EMC Corporation’s views, processes, or methodologies.

Introduction

Microsoft SharePoint offers a wide variety of functionality that includes portal capabilities, collaboration, content management, forms, business intelligence, and search. It has emerged as a major driver in information growth. Organizations of all sizes have rapidly adopted this new collaboration paradigm; an IDC survey shows that over 71% of organizations either currently use or are planning to use SharePoint.¹

SharePoint connects people with content. Besides that it plays another important role; it connects seamlessly with the content creation solution, Microsoft Office 2010 which has more than 100 million customers.² SharePoint and Office integration leads organizations in one direction; huge volumes of content. Volume of content in SharePoint continues to grow exponentially. The IDC Digital Universe study reports that big data is here – by 2020 there will be as much as 35 Zettabytes of information and organizations will be responsible for storing, managing, and protecting 80 percent of the digital universe.³

As Office and SharePoint traction continues to drive forward, we see a lot of organizations leverage it far more now than ever before—in some ways it has become the central point for information access. An enterprise may have thousands of site collections that loosely tie the enterprise content together. This in turn makes SharePoint lack a standard information management strategy that would define enterprise-wide retention, compliance, and records management policies. When organizations require advance content management capabilities and process-centric applications, SharePoint capabilities have to be augmented with partner solutions. In fact, AIIM's Industry Watch survey says over 67% of respondents needed to either customize or buy third-party products and solutions to allow SharePoint to meet their enterprise requirements.⁴ This article discusses EMC's various solution offerings that provide information intelligence to leverage the value of SharePoint at an enterprise level and manage the content across its entire lifecycle—create/capture to archive and retire.

SharePoint Capabilities

Microsoft SharePoint has been widely used to build portals and team collaboration sites to address content access and collaboration among workgroups in the organization. With the release of Microsoft SharePoint Server 2010, SharePoint emerged as a content management solution providing electronic forms management, business intelligence, search, and social networking. SharePoint provides an intuitive user interface that ensures consistent and rich user

experience with Web 2.0 technologies. Seamless integration with Microsoft Office products and relatively easy deployment has fuelled the rapid adoption of SharePoint.

AllM's Industry Watch Survey identifies the top six things SharePoint is being used for:

1. collaboration/workspaces/team sites
2. document management
3. file share replacement
4. web – internal/intranet/staff-facing sites
5. forums, blogs, and wikis
6. portals⁴

A closer look at these six uses points to more content in content repositories. Something to really look at is the third most popular use for SharePoint. It could also mean that not all of the content is business critical, or even being actively leveraged.

SharePoint is not being used for archiving, records management, e-discovery, and external forms processing via capture.

SharePoint Implementation Challenges

SharePoint presents various implementation challenges for enterprise-scale execution. These challenges can be grouped together under three categories:

1. enterprise content management
2. challenges, operational challenges
3. compliance challenges

ECM Challenges

The key limitation of SharePoint from a content lifecycle perspective is that it lacks information capture capability; transforming paper files into business-ready, digital content. The business process management with InfoPath forms has limited capabilities for building composite applications. SharePoint has modest capabilities in imaging and digital assets management.

Operational Challenges

Operational challenges relate to the performance, management, and costs associated with the SharePoint environment. SharePoint recommends certain soft limits for performance reasons such as 200 GB per content database under general usage scenarios. In case of high volume

content management, SharePoint may require more than one repository. Such cases can lead to implementation challenges, from information access to search.

SharePoint requires Microsoft SQL Server to store both content and metadata. As the SQL databases grow bigger, organizations would experience degradation in search and retrieval times. Exponential content growth not only affects performance but also impacts backup operations.

Finally, storage costs should be taken into consideration. SharePoint is a critical ECM application so the larger the content grows, the more organizations are potentially spending on high performance storage.

Compliance Challenges

In most organizations, SharePoint is predominantly used for collaboration. Information shared across collaborative projects might contain sensitive or confidential information. Therefore, IT governance has to be in place for inactive content on how to retire sites and how this sensitive information is controlled and retained, if necessary. SharePoint has no capability for archiving inactive content that must be retained for e-discovery, compliance, or long-term preservation.

This article lists EMC products and technologies to solve these SharePoint implementation challenges.

Intelligent Enterprise Capture

Business critical information comes in many forms—paper, fax, email, or a document in network storage. These are transformed into intelligent content by the use of enterprise content management solutions such as SharePoint. EMC Captiva[®] accommodates single or multiple entry points for capturing documents centrally or from branch offices. It also supports the use of basic desktop scanners to multi-function peripheral (MFP) devices and can pull documents from email messages and network folders. EMC Captiva has tools for automatic classification based on the combination of text- and image-based analysis to identify document types. Automated classification also enables intelligent routing of documents to corresponding lists and document libraries that can trigger a workflow and introduce new templates and content types. SharePoint can thus extend intelligent information capture with EMC Captiva to incorporate scanning, viewing, extracting, and indexing functionalities for an organization.

Captiva transforms paper documents into electronic information that can be stored and managed in a variety of systems. When integrating, Captiva delivers:

- Electronic images, that can be stored in specific locations based on document content or business rules
- Document metadata that can be used to find documents within a large repository
- Extracted data, that can be used to execute business processes
- Trigger Processes, so that Captiva can facilitate fully automated business processes

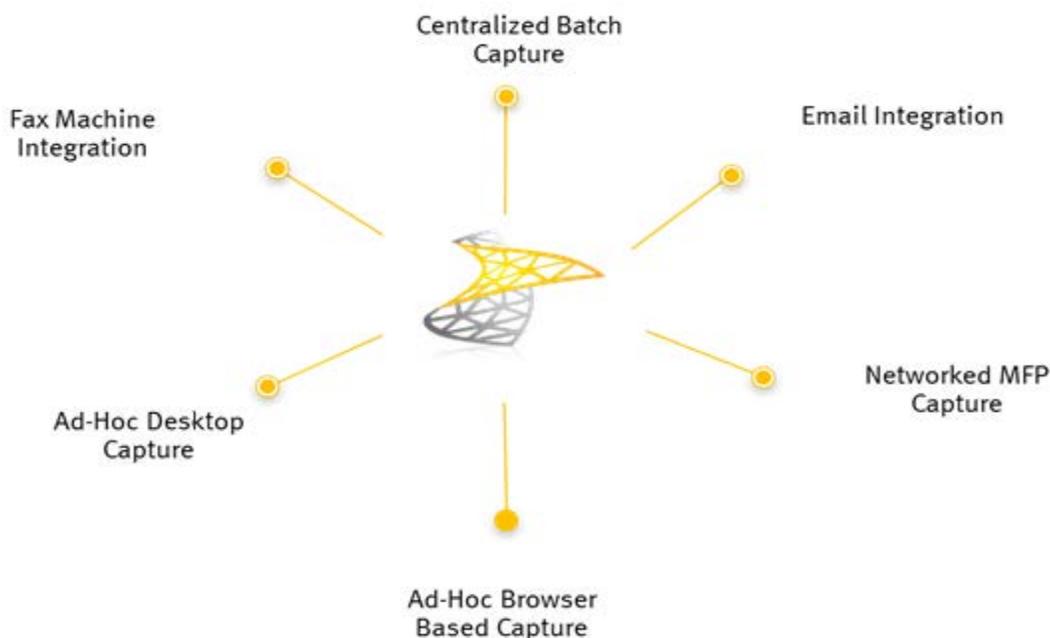


Figure1: Enabling SharePoint for Intelligent Information Capture

Organizations today are using capture solutions in a centralized environment; there is a growing movement among organizations to push the capture of information to the locations where paper first enters business process. Studies have shown distributed capture is expected to grow significantly so that more organizations will look for solutions to automate the capture of paper at remote locations and integrate distributed capture with business processes. This movement to distributed capture can be attributed to a number of factors, including lower bandwidth cost, desktop scanners, MFP devices, and to move information electronically versus shipping paper. Captiva is architected to handle this need for remote indexing in a variety of ways.

Key integration features of EMC Captiva Intelligent Enterprise Capture for SharePoint include:

- Mapping values to properties
- Converting document batch levels to SharePoint folders
- Customizing file names

- Assigning a document profile
- Exposing file formats
- Documenting version control
- Ensuring data integrity

Some of the key reasons and benefits sought by organizations include improved business process, customer service, and reduced document processing time. SharePoint with an intelligent capture solution such as EMC Captiva can handle high volumes of critical documents by incorporating scanning, viewing, image enhancing, indexing, and extracting functionalities to a collaboration and document management platform. This solution processes high volume content, reduces costs, improves information accuracy, and streamlines business process. Thus, EMC Captiva can extend the advance content management capabilities of SharePoint.

Intelligent Information Access

SharePoint goes well beyond basic content management into portal, search, and collaboration technologies. It provides an intuitive user interface and ensures a consistent and rich user experience. Additionally, it provides connectors to other LOB systems.

Comparing SharePoint with Documentum finds that Documentum offers more capabilities than SharePoint. The only vendor to have effectively built a "content management stack" that can manage content across its entire life cycle,⁵ the highly scalable Documentum supplements its core components with a number of archive, process, digital asset, forms, analytics and storage options. Figure 2 depicts the advance content management capabilities of SharePoint and Documentum.

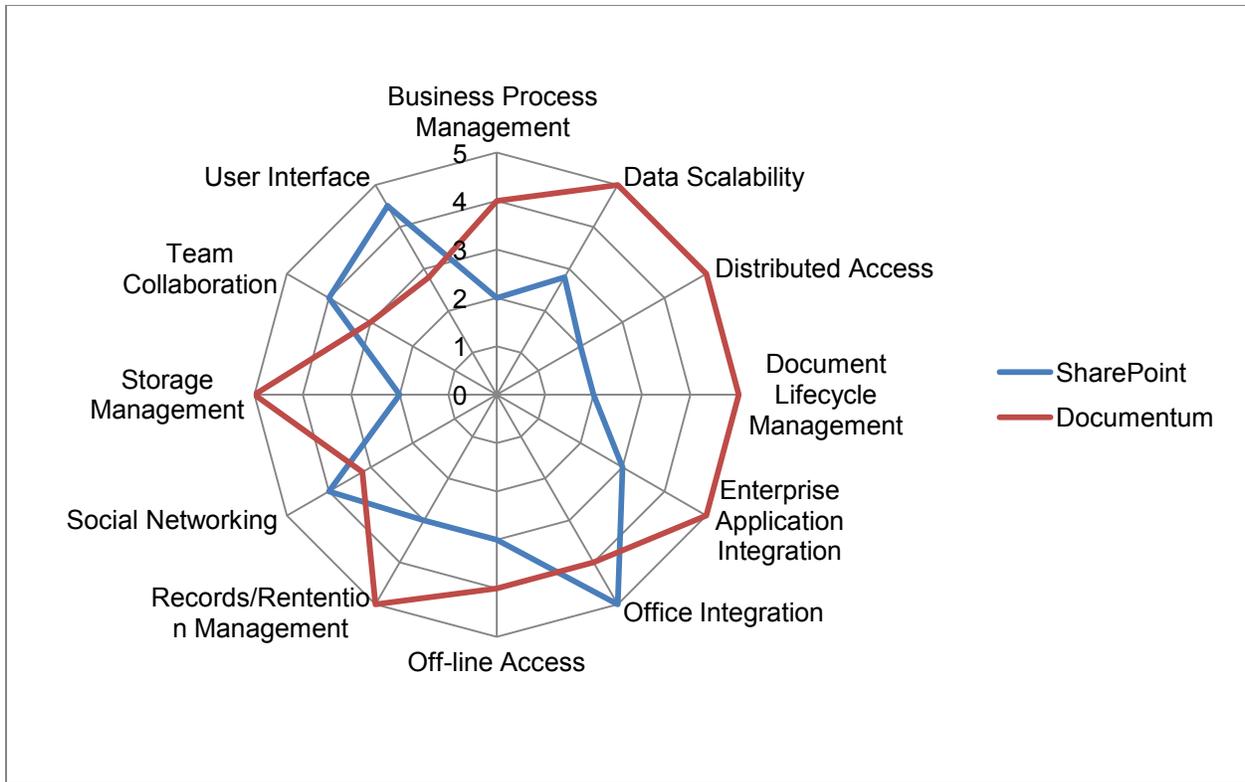


Figure 2: Advance Content Management Capabilities of SharePoint and Documentum

When there is a requirement to involve SharePoint users and content within broader business processes that span the information architecture or when there is a need to extend and enhance SharePoint with additional ECM capabilities, it can be made possible by integrating SharePoint with EMC Documentum. It implements standardized policies and controls to participate in critical business processes or access enterprise content within Documentum through SharePoint interface. Table 1 lists capabilities of SharePoint and Documentum, along with necessary integration requirements.

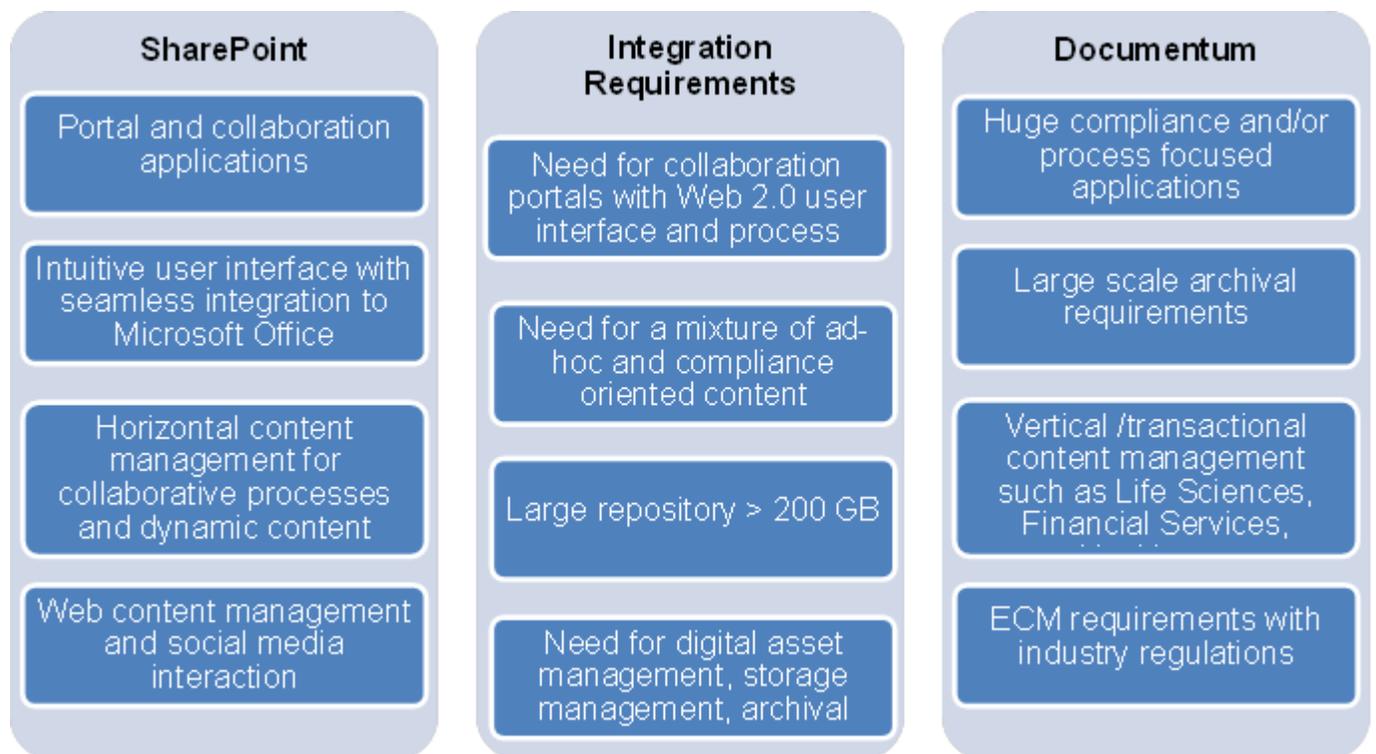


Table 1: SharePoint and Documentum Integration

The various integration connectors for intelligent information access between SharePoint and Documentum are:

Content Management Interoperability Services (CMIS)

CMIS is a web services standard for managing content with any compliant content management system. It was ratified by the Organization for the Advancement of Structured Information Standards (OASIS) in 2010. CMIS focuses on the basic content capabilities of ECM—create, read, write, delete, and query functionalities for organizations with disparate content repositories. CMIS provides both SOAP and REST interfaces and is supported by Microsoft SharePoint Server 2010 and Documentum Content Server 6.0 or higher.

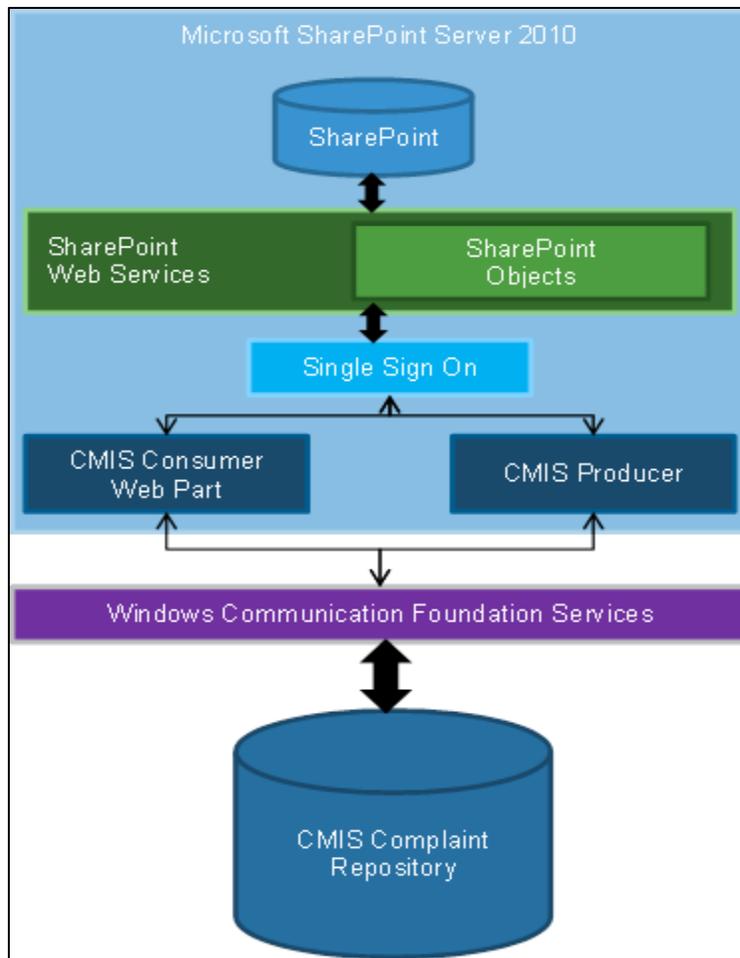


Figure 3: SharePoint CMIS

CMIS connector for Microsoft SharePoint Server 2010 enables SharePoint users to interact with content that is stored in any repository that has implemented the CMIS standard. It also makes SharePoint Server 2010 content available to any application that has implemented the CMIS standard. This connector, available as part of the SharePoint 2010 Administration Toolkit, includes 2 features:

- Content Management Interoperability Services (CMIS) Consumer Web Part
- Content Management Interoperability Services (CMIS) Producer

My Documentum for SharePoint

My Documentum for SharePoint provides quick, out-of-the-box deployment with a set of web parts in SharePoint. These web parts leverage SharePoint single sign-on (SSO) to offer seamless and direct user access to content being managed within a Documentum repository through SharePoint user interface. If SSO has not been configured, users need to log in to Documentum after logging into their SharePoint site via a browser. SharePoint server service

gathers credentials to pass to Documentum. Documentum Foundation Service (DFS) passes user credentials to Content Server to get a session for the SharePoint user and grants access to the Documentum library.

My Documentum for SharePoint allows end-users to search, view, and edit content in Documentum through SharePoint and Microsoft Office. It also exposes certain Documentum-specific functionalities such as virtual documents, lifecycles, and renditions. It also supports critical data management requirements, including regulatory compliance, data retention, and document lifecycle management throughout the enterprise.

The following screenshots show Documentum data accessed via SharePoint.



Figure 4: Documentum Cabinet – Account Management

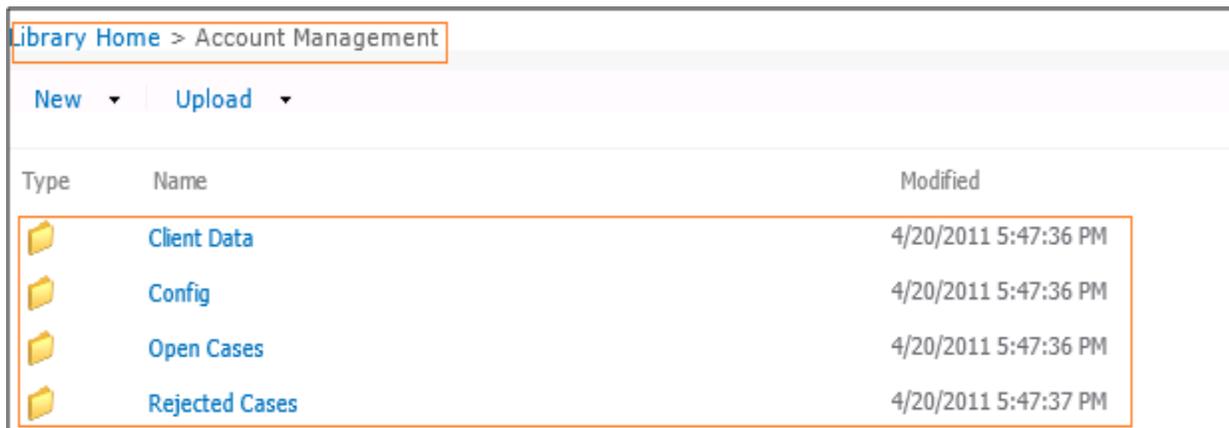


Figure 5: Accessing Documentum Content from SharePoint

However, My Documentum for SharePoint is not transparent. Some Documentum-specific functionalities are exposed in the menus that could confuse an average SharePoint user. An important point to consider is although Documentum functionalities are exposed in SharePoint, My Documentum for SharePoint cannot replace WebTop.

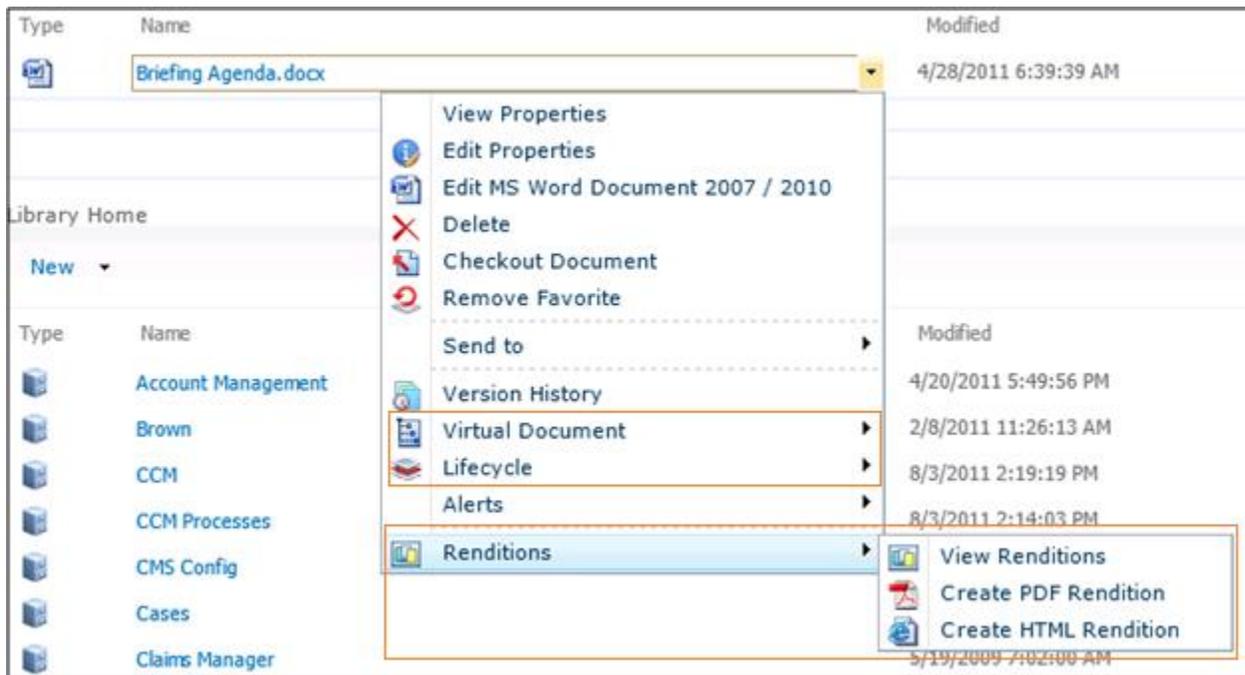


Figure 6: Documentum Functionalities Exposed in SharePoint

My Documentum for SharePoint connects directly to Documentum Content Server and emulates SharePoint user experience for Documentum to provide basic content services. Thus it enables users to contribute to enterprise-wide business processes by accessing content in Documentum without leaving the SharePoint user interface. It is an extensible and flexible approach to bring SharePoint and Documentum together.

EMC Documentum Repository Services for Microsoft SharePoint

SharePoint can be configured to store data within EMC Documentum repository as opposed to its associated SQL Server database content store. EMC External Blob Storage (EBS) handler that extends Microsoft EBS layer brings the content into the Documentum repository. The EBS handler gets involved as data is saved on SharePoint. Content is stored as a Binary Large Object (BLOB) that has the object and that object's metadata. While saving, the metadata is first stored in the associated SQL Server and the object is saved in the EMC Documentum repository. The metadata is also copied from the SharePoint SQL Server to the EMC Documentum repository, where it is joined with its associated object. The metadata does remain

in associated SQL Server to enable SharePoint users to retrieve their content. To ensure consistency and accuracy, if the object's metadata changes, SQL Server updates its copy in the EMC Documentum repository. This process is 100% transparent to SharePoint end users. The following screenshots show how content in SharePoint appears to end users and how the content is stored in Documentum.

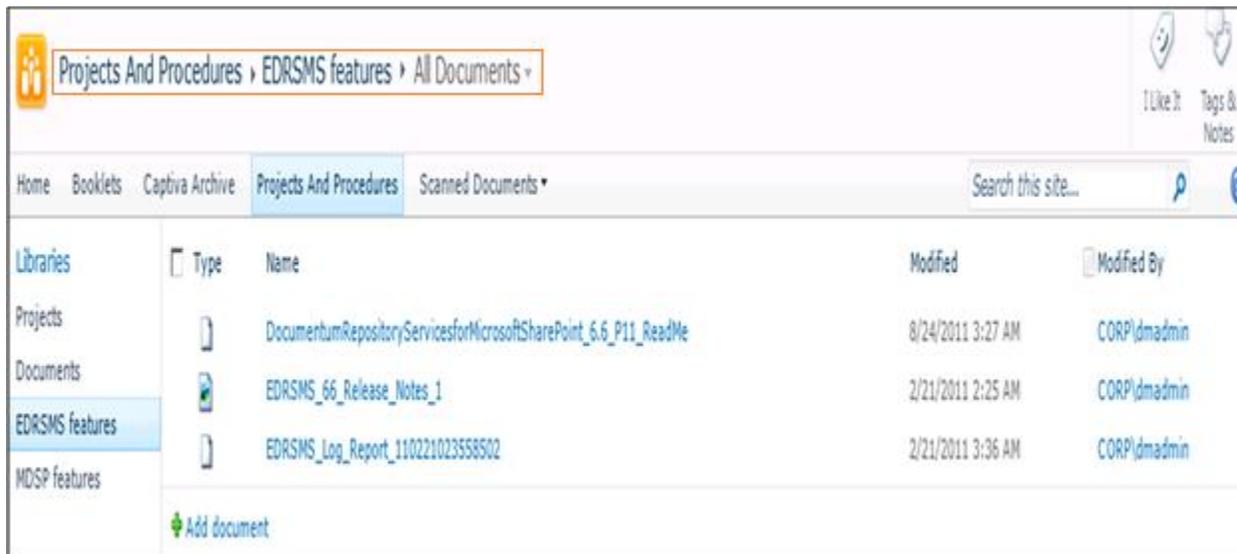


Figure 7: SharePoint Site Transparent to End Users

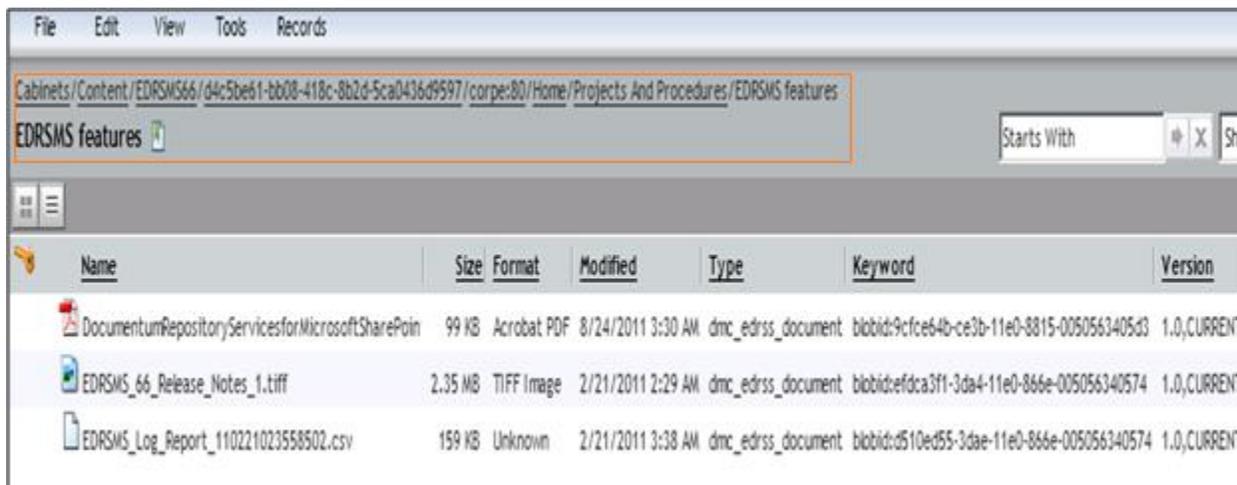


Figure 8: Content Resides in Documentum

When content in SharePoint grows, it is efficient to use SQL Server only for the metadata. Documentum is an important alternative for cost-effective storage. EMC Documentum Repository Services for Microsoft SharePoint externalizes content and provides 100% transparency to SharePoint users and process. It provides compliance needs such as common

policy management, long-term retention, and expunge. This also paves the way to repurpose or reuse SharePoint content in ECM processes. It improves SQL Server manageability and supports deduplication and fragmentation control.

A quick comparison of all three solutions for intelligent information access is shown below:

CMIS	MyDSP	EDRSMS
Allows basic operations on content	Access Documentum content and processes within SharePoint	SharePoint generated content to live within Documentum
SOAP, REST	SharePoint Web Parts	External BLOB Storage
Not transparent	Not transparent	100% transparent
Basic operations	Allows SharePoint as a Documentum client	Extends ECM capabilities of SharePoint
OOB Web Parts, configuration	OOB Web Parts, configuration	Deployment and Configuration

Table 2: Intelligent Information Access Solutions Comparison

Intelligent Information Governance

Many organizations have hundreds of portals and thousands of site collections globally that might contain millions of documents and accounts of terabytes of storage. About 25% of content that resides in SharePoint is inactive content meaning it is either orphaned or has exceeded its shelf-life.⁶ The aged and outdated content is rarely archived, managed for compliance, or considered for long-term preservation. There is no information governance from SharePoint to manage this inactive content. In addition to growing legal and regulatory concerns, inactive content could potentially degrade the performance of SQL Server. Sample information governance in SharePoint is depicted in Figure 8.

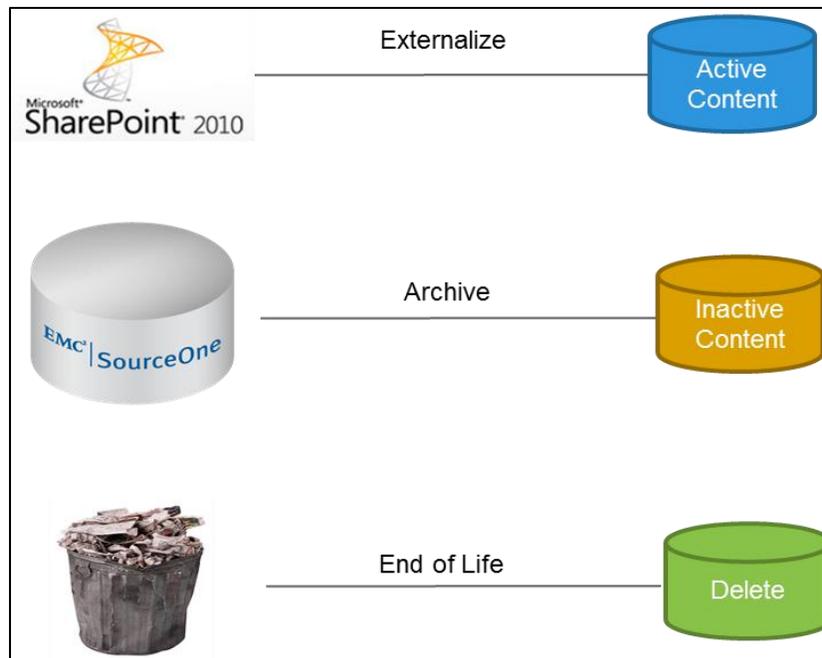


Figure 8: SourceOne to Externalize Active and Inactive Content

Natively, SharePoint stores its content in SQL Server databases for both metadata and the actual content that is stored as Binary Large Object (BLOB). About 80 percent of data for an enterprise-scale deployment of SharePoint Foundation consists of file-based data streams that are stored as BLOB data. Maintaining large quantities of BLOB data in a SQL Server database is a suboptimal use of SQL Server resources. Equal benefit can be achieved at lower cost with equivalent efficiency by using an external data store to contain BLOB data.⁷ EMC SourceOne™ leverage Microsoft’s recommended approach to externalization (RBS, remote BLOB storage) and uses Microsoft API to essentially dissect the content into metadata and BLOB, where BLOB is routed to a SourceOne folder and the metadata to SQL Server database. Key features of EMC SourceOne are:

- **SourceOne RBS Provider** – displaces the content load within SharePoint’s database improving search and retrieval times, and reduces backup and restore times to meet service level agreements
- **SourceOne RBS Scheduler** – externalizes content existing BLOB’s in the SharePoint database to reduce costs and improve performance for both BLOBs that existed before installing SourceOne RBS provider and BLOBs added back to SharePoint due to error conditions
- **SourceOne RBS Health Monitor** – monitors storage locations for low disk space and provides administrators a status overview to ensure optimal storage management and performance

- **SourceOne RBS Migration Tool** – assists the migration of BLOBs from EBS to SourceOne RBS (SharePoint 2007 to 2010) and third-party RBS solutions to SourceOne RBS

EMC SourceOne provides intelligent information governance for SharePoint by:

- Externalizing active content from SQL Server to external, lower-cost, tiered storage
- Archiving inactive content according to regulatory and compliance retention policies
- Reducing the length of backup windows
- Enabling a transparent end-user experience while accessing archived content
- Ensuring litigation readiness by making content readily accessible for e-discovery

The benefits of EMC SourceOne for Microsoft SharePoint solution:

1. Improves the performance of Microsoft SharePoint
 - System performance is in line with service level agreements (SLA)
 - No negative impact to operational tasks such as regular backups and data restore
 - Inactive content is removed from high-availability production storage onto lower-cost tiered storage solutions
2. Enables information governance
 - Retain content according to regulatory or corporate policy
 - Destroy content after meeting all regulatory obligations to retain
 - Easily discover content once a legal request has been made
3. Maintains end-user transparency
 - Allows end-users to search across both native SharePoint libraries and repositories of archived data
 - Leverages the SharePoint user interface
 - Enables administrators to access both active and inactive content

Conclusion

Collaboration with Microsoft SharePoint empowers users, but business realities make them require a new approach to provide a complete enterprise content management solution. An enterprise content management platform requires a reliable, scalable, and cost-effective solution that needs to adopt a best-of-breed approach by leveraging EMC solutions based on specific features or preferences to enable SharePoint to widen its reach across all aspects of enterprise content management. EMC solutions add information intelligence and also ensure cost-savings and performance benefits for mission-critical applications.

Together, Microsoft SharePoint and EMC provide the most compelling enterprise content management platform to create hybrid applications that take advantage of the SharePoint interface and extend advanced content management capabilities of SharePoint. EMC solutions accelerate deployment and simplify SharePoint management, reducing SharePoint site proliferation and corresponding silos by externalizing active and/or inactive content without impacting SharePoint users.

Improve Performance and Reduce Costs

EMC SourceOne for Microsoft SharePoint provides transparent, single point of access and helps with the rapid growth of content in SharePoint through its intelligent, tiered storage management. It can improve performance of active content in production environments and also reduce the operational and management costs associated with inactive content. EMC SourceOne for Microsoft SharePoint shortens backup windows and protects information in SharePoint through low-cost recovery. Total cost of ownership is reduced from the initial capital expenditure to ongoing maintenance and support.

Scalable, Flexible, and Extensible

My Documentum for SharePoint enables users to view and work with content that is stored in Documentum while using SharePoint as the user interface. This brings together the best of both worlds. SharePoint can be used to collaborate content creation. Once the document is complete, information lifecycle management of Documentum can be applied to the content. This is a very effective and flexible approach to content management.

EMC Captiva Intelligent Enterprise Capture for SharePoint is a flexible and scalable enterprise capture solution that includes capture, document classification, data extraction, information validation, and delivery to SharePoint.

Ensure Compliance

EMC SourceOne for Microsoft SharePoint can be used for regulatory compliance, risk mitigation, and eDiscovery. EMC SourceOne also helps in applying full lifecycle management including retention and disposition to inactive content. While inactive content resides outside the expensive production environment, it remains simply searchable through the SharePoint interface. When regulatory and legal requirements of business grow more complex, information in SharePoint can grow exponentially. EMC SourceOne for Microsoft SharePoint will provide the

necessary intelligent information governance for managing content across enterprise information infrastructure.

Safe and secure archiving of SharePoint content within Documentum can also be achieved with Documentum Repository Services for Microsoft SharePoint.

Appendix

1. "Microsoft Office and SharePoint Traction: An Updated Look at Customer Adoption and Future Plans," IDC, October 2009
2. [Microsoft: 400 million Windows 7 and 100 million Office 2010 licenses sold – Partner Conference Statistics](#)
3. [The 2011 Digital Universe Study: Extracting Value from Chaos](#)
4. "AIIM Industry Watch: SharePoint–Strategies and Experiences," July 2010
5. Gartner 2009 ECM Magic Quadrant
6. *InfoTrends Report: "Gathering MOSS?"; August 2009*
7. [External Storage of BLOBs in SharePoint](#)
8. All screenshots are from [EMC Demo Cloud](#)

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED "AS IS." EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.