

# White Paper

---

## **Incorporating SharePoint into Electronic Discovery Processes**

*By Brian Babineau*

**May, 2010**

---

This ESG White Paper was commissioned by FTI Technology and is distributed under license from ESG.

## Contents

Introduction .....	3
SharePoint Proliferation .....	4
Ubiquitous Usage .....	4
Technology Trends are a Catalyst.....	4
Unparalleled Data Growth.....	5
Challenges Abound .....	5
Infrastructure De-centralization.....	5
Organized Chaos.....	6
Architectural Complexity.....	6
Electronic Discovery Looms.....	6
Same Data, Different Place.....	6
Learn from the Past .....	6
What Lies Ahead.....	7
Addressing the Challenges.....	8
Be Prepared: Establish a Plan .....	8
The Bigger Truth .....	10

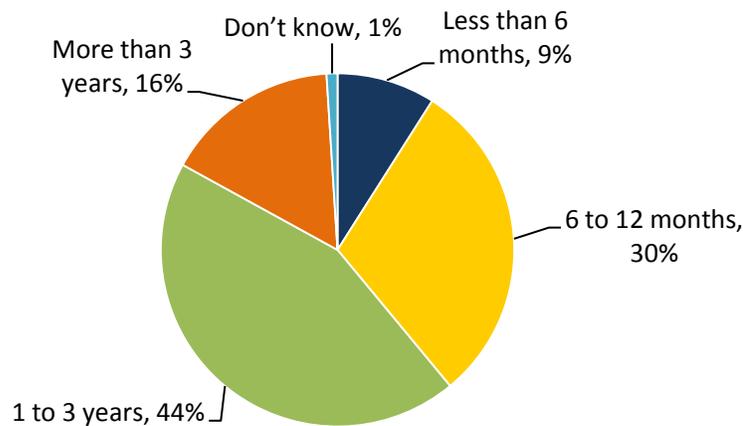
All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change from time to time. This publication is copyrighted by The Enterprise Strategy Group, Inc. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of the Enterprise Strategy Group, Inc., is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact ESG Client Relations at (508) 482-0188.

## Introduction

Microsoft first introduced SharePoint in 2001 as an early entrant in the growing Web portal and content management market. Since then, it has emerged as one of the most popular and widely deployed business applications—and this trend shows no signs of abating given Microsoft’s focus (SharePoint is a key competitive weapon in the fight against Google Apps) and maturing customer adoption. A 2009 ESG research study found that while most SharePoint deployments were in their infancy (i.e., just being transitioned from development to live environments, see Figure 1), senior executives still claimed that the technology was going to be a top ten IT initiative in 2009 and 2010.<sup>1</sup>

*Figure 1. SharePoint Deployment Status in 2009*

**For how long has your organization been using Microsoft SharePoint in a production environment? (Percent of respondents, N=263)**



*Source: Enterprise Strategy Group, 2009.*

The growing expansion of SharePoint is happening in parallel with a shift in legal and regulatory processes. More inquiries and subsequent matters now include the discovery of electronically stored information (ESI), or “eDiscovery.” Just when corporate counsel started to formalize electronic discovery processes around common ESI sources such as desktops, e-mail, and file shares, SharePoint arrives on the scene to further complicate things.

Although there hasn’t been any precedent established regarding the discovery of information in SharePoint, corporate counsels should prepare themselves for several reasons. Much of the relevant data being produced today may very well be stored in SharePoint in the future as the application replaces network file shares. Desktop virtualization proliferation, by which no information is stored on physical PCs, means that users are likely to save information in centralized websites like SharePoint. Microsoft is also making it easy to copy messages from Exchange environments into SharePoint, where they can be managed according to business policies and predefined workflows. Lastly, integration between Microsoft Office 2010 and SharePoint will make it very easy for users to save content within the latter, making it easy to bypass desktop drives and corporate file shares. The bottom line is that regulators and courts are more concerned about whether relevant content is produced rather than where it is stored.

This paper details the aforementioned trends in technology that will drive more content to SharePoint, encouraging corporate counsel to start assembling a formal eDiscovery plan for this data source. SharePoint is unlike any other corporate collaboration source today; its complex, multi-tier architecture makes it hard to identify, collect, preserve, and process content. Its many configuration options require corporate counsel and IT to make choices regarding how to execute these tasks, further substantiating the need for an eDiscovery plan before it becomes a

<sup>1</sup> Source: ESG Research Report, [Microsoft SharePoint Adoption, Market Drivers, & IT Impact](#), March 2009. All ESG research data comes from this report.

mainstream ESI source. ESG can offer some key considerations for an eDiscovery plan as well as justification for how such an effort can pay dividends in the future. Other productivity applications and associated repositories, including SharePoint, consumed via the “cloud” are on the horizon and will, too, need to be added to the ESI data source list.

## SharePoint Proliferation

### Ubiquitous Usage

The aforementioned ESG study revealed that approximately one in two global organizations had already deployed or were planning to deploy SharePoint and 52% of IT managers said it was going to be a top ten initiative over the upcoming 24 months.<sup>2</sup> The reason for this widespread adoption and subsequent prioritization is SharePoint’s ability to centralize collaboration via the Web and its versatility to support many work patterns and tasks. To a knowledge worker, SharePoint is a website where content can be edited, versioned, and seamlessly shared—replacing overcrowded network file shares. To a business process owner, SharePoint is a website where policies can be translated into workflows that are straightforward to enforce. To a marketing executive, SharePoint is a publishing tool that ensures customer- and partner-facing websites always have the latest product information.

Many business users are familiar with SharePoint because it is not burdensome to try. Microsoft Office SharePoint Server 2007’s list price is under \$5,000, with Client Access Licenses under \$100 per seat—these estimates do not account for any volume licensing agreements an organization may have in place with Microsoft. Secondly, once IT sets up a SharePoint “farm” (which includes the application, role, and database servers), it is easy to add new sites. After a site is set up, the administrator can configure the security permissions as well as the workflows that will govern the content stored within that particular site. Once users experience SharePoint, its simplicity breeds popularity; more sites are set up and the cycle repeats.

### Technology Trends are a Catalyst

Most early SharePoint deployments replaced cluttered corporate file shares and, to date, many departments are only using a portion of the application’s functionality—specifically, its version control capabilities. Some may believe that this will be the extent of SharePoint’s usage, but several key technology trends are likely to broaden the way SharePoint is utilized, including:

- **Microsoft Office 2010.** Due out in the middle of calendar year 2010, this is the first version of Microsoft Office that enables individual applications (Word, Excel, etc.) to be delivered online or as a service. The new version of Microsoft’s flagship productivity suite also makes it easier for users to store data within a SharePoint site.
- **Desktop virtualization.** This emerging technology solution is the beginning of thin computing where applications and data are not stored on a physical PC. Instead, applications are accessed and the associated data is stored in a central location and accessed via the Internet. SharePoint becomes a logical place for users to save application data.
- **Business process automation.** Whether it is managing a sales contract or publishing a Web page, the creation of and interaction with almost all content is a process. Editing, approval cycles, and other tasks have to be created and managed. Security permissions—specifically, who can access the content—need to be enforced. Right now, many of these tasks are managed via e-mail, within file shares, or as part of custom applications that are already difficult to manage. Organizations can centralize these operations and supporting tasks within SharePoint.

Captive data is the biggest hindrance to effective collaboration and more automated business processes. The above technology trends are designed to address these issues and SharePoint is a key enabler. This is why ESG believes that, if it hasn’t already, SharePoint will become part of the corporate DNA.

---

<sup>2</sup> Source: <http://office.microsoft.com/en-us/sharepointserver/FX102176831033.aspx>.

## Unparalleled Data Growth

If SharePoint is being deployed to replace file shares, the information shift is rather obvious: when organizations upgrade to Office 2010, evaluate desktop virtualization, or both, much of today’s unstructured corporate information will move from desktop hard drives to SharePoint sites, creating an information explosion within this relatively new application.

SharePoint information growth will also be driven by SharePoint itself because it can serve as an application framework or as a productivity tool. As an example, corporate intranets are likely to be built on the platform so employees are aware of the latest news, events, benefits, and policies. SharePoint also has its own Wiki and blog development capabilities, so employees can create content directly in a site.

## Challenges Abound

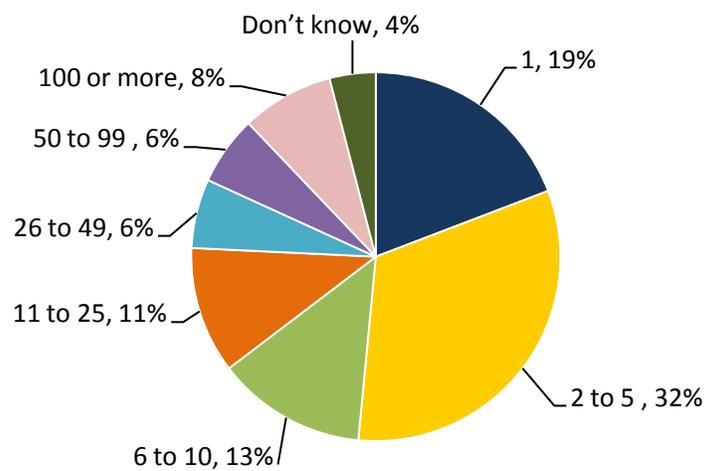
### Infrastructure De-centralization

From both an end-user and a business perspective, the core attraction of SharePoint is the centralization of information. From an IT perspective, however, the view is often very different. All SharePoint content is stored within a SQL Server database and as the database gets larger, response times slow and backups are tougher to complete. As a result, in SharePoint Server 2007 and 2010 environments, Microsoft recommends that the underlying SQL Server database be capped at 100 GB and 200 GB, respectfully. For many organizations, this threshold is triggered within days of configuring the first site.

To avoid the SQL Server database limitation, IT often splits SharePoint back-end servers. A single farm is likely to consist of several SQL Server databases or IT will set up new farms, each with its own content database. SharePoint infrastructure sprawl worsens when farms are deployed at remote sites to satisfy mobile workforce requirements. According to ESG’s study, over 75% of organizations running SharePoint in 2009 had farms deployed in multiple physical locations (see Figure 2).

*Figure 2. SharePoint Typically Runs in Multiple Physical Locations*

**To the best of your knowledge, in how many physical locations is SharePoint currently deployed across your entire organization?  
(Percent of respondents, N=339)**



Source: Enterprise Strategy Group, 2009.

## Organized Chaos

While it is possible to configure SharePoint to provide visibility and management across multiple farms, uptake often happens so fast that organizations might not have the luxury of far-sighted implementation planning. IT usually ends up with control over the technology infrastructure, but business users are empowered to set up sites, lists, and workflows and determine security parameters. Rarely does IT know what a site is being used for or what content is in it. This also means that IT is often unaware of orphaned sites or stale content.

There are so many challenges with SharePoint when it is handled by IT, one can only imagine what happens when it isn't. SharePoint's footprint is widened further by "non-sanctioned" implementations that arise as users create sites to meet ad hoc business requirements. These deployments are not part of the corporate data topology and can be located anywhere. These types of implementations are more common than many in IT might think; some departments fund their own SharePoint projects prior to IT adding the application to an internal service catalog.

## Architectural Complexity

SharePoint is a multi-tier application involving Web, role (search, etc.), and database (content) servers along with various other infrastructure components in addition to the database servers where enterprise content is stored. Most organizations leverage existing Microsoft Active Directory instances to manage security across one or more SharePoint farms.

Depending on security permissions, a user may be allowed to create a site or simply read content on a site. Within a site, there may be multiple document lists and Web parts, all of which inherit the security permissions established for the site. Certain files saved within a site may trigger a workflow that includes document forwarding and approvals, which are tracked. The number of configuration permutations within a given SharePoint implementation is great, making it extremely flexible to address almost any business collaboration problem within any given department.

When SharePoint stores content, it does so differently from traditional corporate repositories. It does not store "files," it saves BLOBs (Binary Large Objects) consisting of metadata and the actual file. The metadata includes the security permissions, owner, and creation date of the file and is used by SharePoint to manage the information (i.e., enforce a specific workflow for all spreadsheets saved by the finance department).

## Electronic Discovery Looms

### Same Data, Different Place

It is hard to say that legal departments are "comfortable" with electronic discovery right now, but most are, at a minimum, aware of what IT departments go through to identify, collect, and preserve ESI from desktops, e-mail applications, and file shares. SharePoint now must be added to the list of potential ESI sources for any given matter because it is storing relevant data being generated in more common formats (loose productivity files such as spreadsheets, PDFs, contracts, etc.) and emerging content types (wikis, blogs, etc.). For corporate counsels to think that such inclusion is anything but a foregone conclusion is myopic because SharePoint is storing the same information being "discovered" today in the form of loose files and e-mail. In addition, SharePoint is being used to store new formats of information such as wikis and blogs.

### Learn from the Past

If organizations wait for a matter that includes SharePoint before they put together a plan, they are going to be caught in the middle of an extremely unpredictable (in terms of cost and time) situation. Companies should not forget what they went through the first time they had to discover e-mails where processes spanned the primary messaging environment, backup tapes, desktops, and file shares (where personal mail archives are frequently stored). Most corporate counsels were so caught off guard by the time and effort involved that they either settled or tried to argue e-mail out of scope. After going through a painful and expensive process of discovering a "new"

ESI source once, most organizations determine that they need a more formal, automated process. Corporate counsels have the opportunity to implement such a process with SharePoint—a luxury that should be taken advantage of.

## What Lies Ahead

The electronic discovery process is not going to change every time a new data source appears. However, this process must be analyzed in the context of the data source. Whether determining identification methodologies or preservation policies, e-mail introduced new issues when compared to desktop data—and SharePoint will surely present its own unique challenges.

- During **identification**, organizations do have the option of using SharePoint’s search capabilities. This must first be licensed and turned on for every SharePoint farm an organization needs to query. The good news is that results can be federated across any number of SharePoint sites so long as search has been activated. If a site has to be included and doesn’t have search enabled, IT can enable the capability. This triggers an indexing operation, which, depending on the number of documents within a site, can take time. SharePoint’s search capabilities also force IT and legal into a “keyword-based” identification approach. While this may be suitable for some matters, there are other situations where internal terminology, frequent misspellings, or synonyms cannot be accounted for. Also, many matters center on a specific set of custodians or employees and attorneys only want to execute searches against these individuals’ repositories including SharePoint. As such, attorneys should fully understand some of the limitations of using keyword searches only. Secondly, ESG advises corporate counsel to, at a minimum know what the policy is for SharePoint search licensing and if all sites are indexed when this licensed is enabled in case they do execute keyword searches against SharePoint. Failure to do so may lead to missing certain documents during this critical first step in the electronic discovery process.
- After identifying relevant content, it must be **collected and preserved** in a forensically sound manner. SharePoint utilizes multiple tiers of metadata, including system-level (site information, list inclusions, etc.) and content-level metadata (identity of owners and editors, dates of access and modification, etc.). For those looking for an analogy, this is the equivalent of preserving all e-mail header information, the message, the attachment metadata, and the attachment itself. Companies must have the appropriate measures in place to collect all SharePoint content inclusive of all the metadata and context.

The next decision companies face, given SharePoint’s workflow and version tracking capabilities, is determining what versions of documents to actually collect. If the versioning capability is turned on, there are likely multiple iterations of the saved document within a site. When a workflow is in use, a document may be at a certain stage, but not final or approved. For those unfamiliar with SharePoint, the decision regarding what versions to collect may be a surprise as a traditional user doesn’t always “see” all iterations of a document even though they are being stored in the underlying SQL Server content database. Corporate counsel should inquire about the different workflows supported by SharePoint as well as if versioning is turned on. It will also be helpful to know how many versions are supported within a SharePoint site so they can determine the volumes that may be collected if they choose to collect and preserve all versions or just a portion.

When it comes to preservation, corporate counsel will need to determine how to preserve the information. SharePoint is a live application which end-users access on a regular basis. Microsoft does offer retention management capabilities which can be set at the site, list, or document level. However, corporate counsel may also want to copy the relevant information from a site to a separate legal “repository of record” where it can be centralized with other data (e-mails, files, desktop images, etc.). The “copy to preserve” process is very common for other data sources and can be applied to SharePoint, but as mentioned, all of the metadata must be included in this operation.

- After gathering the data, it is typically sent for **processing**. Sometimes, there is an initial processing phase executed within the corporation, allowing attorneys to complete a **first pass review**. Following the first

pass review, data is often sent to a legal service provider or law firm for additional **processing**, numerous tiers of **review**, and **production** (note: in many instances, the data is sent directly to the legal service provider for processing). Whether they have chosen to process collections externally, internally, or both, organizations need to choose their processing solution carefully when SharePoint data is involved. Not all processing solutions are capable of handling SharePoint data and metadata appropriately, which means that certain files may not be included in the review process.

Organizations should understand if their onsite or outsourced solutions can or have plans to support SharePoint. Ideally, companies want to avoid situations in which separate solutions are used to process SharePoint and non-SharePoint information as this can increase the probability of inconsistent review.

Getting the initial phases of eDiscovery right is critical in order to ensure complete identification and collection is executed. Organizations must do as much as they can to address their “duty to preserve” requirements. They also want to centralize review processes, reducing the cost as well as the risk of improper or inconsistent coding, which may result in the improper sharing of privileged information. More importantly, improper management of these initial eDiscovery tasks increases risk and cost while decreasing the likelihood of a favorable case outcome.

## Addressing the Challenges

### Be Prepared: Establish a Plan

Everyone knows that being prepared for any situation is better than the alternative. There are countless examples of this unpreparedness pertaining to electronic discovery: organizations spending millions of dollars, scrambling to produce messages and attachments; a company engaged in a case realizing how cumbersome and expensive e-mail eDiscovery was going to be and deciding to settle; and ongoing matters where companies simply couldn’t properly execute eDiscovery of e-mails and incurred substantial monetary penalties.

The complexities of SharePoint, combined with its growing usage, indicate that when it comes to electronic discovery, SharePoint could be worse than e-mail. Companies can take a “wait and see” approach and then deal with consequences and costs just like they did with e-mail—or they can be proactive. The latter requires a level of commitment from the legal and IT departments, but is feasible. A successful project will include:

- 1) **Establishing a SharePoint data map.** Organizations should have basic data maps of critical administrative systems. SharePoint—at least sites that support business processes that create and maintain business records such as contracts, product information, patent applications, etc.—is a logical addition to these data maps. Ideally, IT would have a more detailed understanding of sanctioned SharePoint environments: where SharePoint farms are deployed, what types of data are being produced by which groups, and where data is stored. It is also important to obtain knowledge about SharePoint configurations as well, including the licensing and use of search functions and what kinds of versioning policies are in place. Lastly, there should be a corporate policy regarding the establishment of SharePoint sites, reducing the number of “rogue” sites that IT and legal may not know about.
- 2) **Create standard, repeatable processes.** It may sound generic, but keep in mind that most organizations do not currently have any SharePoint-specific electronic discovery processes in place right now. Establishing these processes involves making some basic decisions such as the best way to collect data to exporting it out of SharePoint for preservation purposes. Companies might investigate:
  - **Custodian-based identification.** When e-discovery requests occur in SharePoint environments, a custodian-based collection strategy is often most effective. This approach enables IT to first determine what documents a user (and by “user” we mean a login credential, as a user can have several of these) has access to. Then, searches can be run against the documents which the user has access to identify what content the user has authored, created, edited, read, or otherwise interacted with. In contrast, a keyword-based approach means that every SharePoint site has to be queried to execute a comprehensive identification process—a process that can be extremely time-consuming depending on how many SharePoint documents a user may have authored or had

access to. In addition, starting with a keyword search against multiple SharePoint sites may return inconsistent results depending on how frequently these sites are indexed, the weighting of the indexes, and if the search engine, whether Microsoft's or a third-party's, returns exact results and proximity results that include misspellings or derivatives of the word.

- **Multiple point-in-time collections.** SharePoint data is, by its definition, dynamic—constantly being added to, updated, and deleted by users and groups. For cases that can drag out over long periods of time, collecting dynamic data can be difficult and can disrupt production cycles as employees are forced to change behavior to accommodate the e-discovery process. Collection activities can be optimized by taking a point-in-time image of SharePoint and then running subsequent point-in-time collections as a case requires. These collections are based on the criteria set forth in the initial custodian-centered collection or any additional parameters established as the matter progresses through the Meet & Confer and initial argument phases.

Incremental point-in-time collections provide the ability to track changes from one instance of data to the next. The nature of each case and specifications of the discovery request will determine the appropriate number and frequency of images that will be required. Some cases, such as those involving contract disputes, might require as much data and as many versions as possible. For other matters, such as insider trading cases, a single point in time view of SharePoint documents might be sufficient as attorneys might focus more on e-mail and other communications data.

- **Comprehensive data exports.** After finding the right data, it needs to be extracted from the SharePoint environment before it can be deleted or modified. This means extracting all metadata along with the core files—a process that is not that easy given how SharePoint stores data. The metadata is crucial to supplying context for the evidence; providing proof of the individuals involved in its production and modification, relevant timelines and workflows; and the setting in which activities took place. This step is extremely difficult to execute and may require specialized IT skill sets (individuals with SharePoint architecture knowledge) and purpose-built tools that provide an audit trail of the actual copy/export activities. Organizations should identify SharePoint experts along with the appropriate supporting technology solutions as part of the overall plan (as opposed to doing this during the matter).
- 3) **Conducting a “trial” run.** Legal and IT departments are extremely busy and do not need to add more tasks to their workload—especially “fake” ones. But it is worth taking an old, closed matter to see how it would have been managed if SharePoint would have been considered in-scope. This would give companies the opportunity to see if they can rely on a data map, introduce a chance to communicate with service providers to see if they can process SharePoint data, and see where IT and legal may not be on the same page.
  - 4) **Optimizing different stages.** Given the vast amounts of information being managed by SharePoint, it is unlikely that all phases of the electronic discovery process can be handled by native SharePoint tools (search) or via IT staff (manual scripts, etc.). There will be opportunities to invest in technology or extend existing eDiscovery solutions to support SharePoint eDiscovery activities just as many have done for desktops and e-mail. There will also be opportunities to leverage legal expertise to validate certain processes to ensure policies and procedures are defensible.
  - 5) **Maintaining flexibility.** SharePoint is not the only trend in document productivity applications. Many organizations are currently evaluating running SharePoint in the “cloud” via Microsoft Business Productivity Office Suite (BPOS). Such an implementation means that the actual data is stored offsite at a service provider. Companies planning to take advantage of this model need to ensure that legal and IT teams are prepared to extend electronic discovery processes to SharePoint and then to BPOS. This requires a plan in place for the former first—one that can be modified as necessary as organizations leverage cloud computing strategies.

## The Bigger Truth

Although the central theme of this paper is the intersection of SharePoint and electronic discovery, a large portion of the discussion is dedicated to the technological intricacies of the former. Legal and IT departments need to be aware that SharePoint is not just another ESI data source; it is a very complex application that facilitates real time information creation and collaboration.

SharePoint adoption is fueled by several technology trends, but companies need to realize that the information being created and saved within it is often the data that is being requested in legal and regulatory matters today. To date, there haven't been any court opinions issues on handling with SharePoint, but there are roughly a handful of cases cited where companies have been forced to produce SharePoint information. ESG believes that is still the "calm before the storm," presenting a unique opportunity for companies to be proactive and put a SharePoint electronic discovery plan in place before the first request arrives. Doing so would alleviate some of the cost and risk of having to figure it out on the fly—something that companies are all too familiar with when it comes to electronic discovery. If this isn't motivation enough, organizations must be aware that SharePoint is just the beginning of communication and collaboration transition. Cloud-based SharePoint is coming soon.



Enterprise Strategy Group | **Getting to the bigger truth.**