BNA INSIGHT

John P. Collins, of The Ingersoll Law Firm LLC, provides an overview of Microsoft Share-Point (SharePoint) to enable legal professionals to gain a better understanding of the discovery obligations which may arise for information residing in SharePoint—and how to start addressing those obligations.

Everything You Need to Know About Sharepoint But Didn't Know How to Ask

By John P. Collins

This article is an introduction to Microsoft Share-Point for legal professionals. Because of the breadth and complexity of SharePoint, it addresses what legal professionals should know at a bare minimum to identify, preserve, and collect electronically stored information (ESI) from the platform, focusing especially on identification.

What is SharePoint?

SharePoint is a technology ''platform'' (as opposed to a single purpose application such as Microsoft Word) which provides a wide array of collaboration, web presentation, content management, information management, and business process related applications.

SharePoint consists of a family of web-centric technologies, features, and functions—all of which are highly customizable and flexible and can be deployed and used in a number of ways.

To better understand SharePoint, it's helpful to recognize it has the capacity to function in various roles.

Generally, there are seven roles in which SharePoint is frequently deployed.[1] For legal professionals responsible for conducting discovery of data residing in Share-Point, the question of what SharePoint is really devolves to a question of how SharePoint is being used. How SharePoint is used shapes the nature and types of data it might produce for discovery.

Seven SharePoint Roles

(1) Collaboration. SharePoint enables the creation of ''Team Sites,''[2] which are web based collaboration sites that allow constituents from both inside and outside an organization to work together, such as on a specific

---

[1] These are not the only roles or ways SharePoint may be used. For a broad discussion about how SharePoint may be used, see **How to Do Everything**™**: Microsoft**® **SharePoint**® **2010** by Stephen Cawood, McGraw-Hill, 2010.

[2] Team Site is the SharePoint term for a basic collaboration site in SharePoint. Technically, Team Site is a template which provides standard collaboration features such as shared calendar, tasks, document library, announcements, and discussions.

project or to facilitate regular inter-departmental information sharing. Team Site functions include document and file sharing, group calendars, discussion threads, blogging, wiki's, etc. Collaboration is perhaps the most widely used SharePoint function, and what many individuals think of first when referring to SharePoint.

(2) Intranet. SharePoint is a platform with the requisite tools to enable an organization to build their corporate intranet.

(3) Extranet. Extranets allow partners and customers to access and interact with information stored ''behind'' an organization's firewall.

(4) Portal. The concept of a portal is to aggregate multiple resources onto a single web site. It is useful to think of a portal as a launching pad or doorway to other resources. The term portal is sometimes used as a synonym for intranet or extranet.[3]

(5) Public web sites. SharePoint is a platform on which an organization may build and publish its public web presence (for example, www.microsoft.com).

(6) Application development and delivery. SharePoint may be used to create business processes, workflows, and applications, to meet business needs. For example, using SharePoint ''out of the box'' an organization could build a contract management system utilizing workflows that automatically notify designated employees certain contracts are about to expire, and who to contact to renew or cancel the contract.

(7) Enterprise Content Management (ECM) platform. ECM is the concept of managing the entire lifecycle (creation to disposition) of unstructured data. There are a number of ECM products available in the market, and SharePoint competes with these products (a full discussion of SharePoint as an ECM suite is beyond this article's scope.) SharePoint provides ECM functionality in three main areas: document management, records management, and web content management.

* * * * * *

These seven roles are not mutually exclusive. An organization might use SharePoint as an intranet, portal, and ECM platform, or just to deliver extranet functionality. ECM may be considered complimentary to all of SharePoint's roles and uses, as its purpose is to manage the creation and disposition of data arising out of SharePoint's use.

---

[3] In the SharePoint universe the terms portal, intranet, and extranet are sometimes used interchangeably. However, some SharePoint literature recognize portal functionality to be a super-set of functionality available only in SharePoint Server 2010.

> **"The complexities of SharePoint, combined with its growing usage, indicate that when it comes to electronic discovery, SharePoint could be worse than e-mail."**
>
> BRIAN BABINEAU
> *INCORPORATING SHAREPOINT INTO ELECTRONIC DISCOVERY PROCESSES*
> © 2010, ENTERPRISE STRATEGY GROUP, INC.

What are the Primary Technical Elements of SharePoint?

While somewhat of an over-simplification, from a technical perspective SharePoint may be thought of as having two key elements: a front-end and a back-end.

Front-end. These elements enable users to access and interact with SharePoint:

■ **Web browser**—the various pages, lists, document libraries, calendars, wiki's, blogs, workflows, applications etc. are accessed by users via a web browser.

■ **SharePoint Workspace**—a software application that is installed on a user's desktop or laptop computer, and enables users to work with SharePoint data offline. SharePoint Workspace is included in some Microsoft Office packages.

■ **Microsoft Office 2010**—there is seamless integration between SharePoint and Microsoft Office. For example, a Word document can be checked in and out of SharePoint directly from within Word, and calendar and contact entries in SharePoint can be entered, edited, and viewed from Outlook.

■ **Microsoft Office SharePoint Mobile**—enables users to access and take SharePoint data offline to their mobile device if they are running Windows Mobile 6.5 or later.

Back-end. The so-call back-end consists of servers (application, web, and database) on which SharePoint runs. These servers are organized as ''farms,'' and can consist of a single physical server or multiple servers; in some large installations, there can be multiple server farms.

Each server has a specific role, as follows:

■ **SharePoint application server**—provides the logic and functional capabilities of SharePoint

■ **Microsoft Internet Information Services (IIS) server**—renders SharePoint content and functions as a web page—it takes the data stored in the database and enables it to be viewed via a web browser

■ **Microsoft SQL Server database**—a relational database on which *all SharePoint content* is stored, whether it's a Word document, Excel spreadsheet, calendar item, blog entry, wiki page, audio file, graphic, etc.

SharePoint Versions

One of the first questions to ask when tackling discovery from SharePoint is "What version is in use"?[4]

The answer impacts how discovery from SharePoint should proceed because the features and functions of SharePoint and the location, retention, distribution, and disposal of its data is predicated in part on the SharePoint version in use.

Given the continued growing interest in cloud computing, when discussing versions of SharePoint, it's prudent to ask whether SharePoint is deployed as an "on-premise" or SaaS[5] (cloud based) system.

(1) Is the organization using an on-premises or hosted version of SharePoint? "On-premises" means the SharePoint servers and other infrastructure reside within the organization's own data centers, and management of those servers and infrastructure is directly under the organization's control.

"Hosted" means SharePoint is provided on a subscription basis by a service provider (for example, Microsoft or Network Solutions.) In the hosted scenario, the IT infrastructure resides in the provider's data center, and basic server administration is handled by the provider.[6]

(2) What version is in use? Multiple versions of SharePoint have been released over the years, and the features and functions that are available to create, store, manage, and purge data is determined in part by the version in use.[7]

The most recent versions of SharePoint available are:

■ **Microsoft SharePoint Foundation.** This version of SharePoint is included, at no additional charge, with the purchase of Microsoft Windows Server 2008, and includes the ability to provision Team Sites for collaboration, extranets, intranets, portals, basic workflows, and other functions.

■ **Microsoft SharePoint Server 2010 Standard.** This version of SharePoint requires additional licensing fees. All SharePoint Foundation functionality is included. Significant additional features include: more power search functions, social networking, records management, legal hold, and ediscovery.

■ **Microsoft SharePoint Server 2010 Enterprise.** This version of SharePoint also requires additional licensing fees. All the functionality available from SharePoint Foundation and SharePoint Server 2010 Standard is included, plus additional business process and integration capabilities such as Excel Services, InfoPath forms services, and Performance Point Services.[8]

SharePoint Ediscovery Considerations

SharePoint is complex from a conceptual, technical, and ediscovery perspective.

The focus up to this point has been de-mystifying SharePoint. This section identifies several key dynamics legal professionals should consider when faced with identifying, preserving, and collecting data from SharePoint.

(1) Identify what version of SharePoint is deployed. As discussed earlier, whether an organization uses on-premises or hosted SharePoint impacts who has actual direct control over certain aspects of the IT infrastructure in which SharePoint data resides.

Further, whether an organization is using SharePoint Foundation or SharePoint Server 2010 determines the features and functions available to create, manage, distribute, and dispose of data.

Determining as early as possible the version of SharePoint in use is essential for creating an effective SharePoint discovery plan.

(2) Identify how the organization uses SharePoint. There are, as mentioned earlier, seven readily identifiable roles in which SharePoint may be deployed, and the types of content and data available for discovery will reflect these roles. For example, a public website is going to contain information acceptable for widespread dissemination, such as investor relations data, product and services brochures, career opportunity listings, etc.

A public web site is not likely to contain proprietary data or intellectual property. However, an extranet may very well contain proprietary data and intellectual property.

Collaboration sites (Team Sites) will contain information representative of the site's purpose—marketing data for a marketing oriented site, project plans and documents (CAD drawings for example) for a project site.

---

[4] Stating an organization "uses SharePoint" is like saying an organization "uses Microsoft." Knowing an organization uses SharePoint does not explain how the organization uses it or what version is in use.

[5] "Software as a service (SaaS, typically pronounced [sæs]), sometimes referred to as "on-demand software," is a software delivery model in which software and its associated data are hosted centrally (typically in the (Internet) cloud) and are typically accessed by users using a thin client, normally using a web browser over the Internet." From Wikipedia at http://en.wikipedia.org/wiki/SaaS

[6] If SharePoint is hosted, who is the provider? While Microsoft is making a significant push in the market with Office 365, and cloud computing is a hot topic, in reality SharePoint hosting has been around for a number of years. Legal professionals should determine who the vendor is and what types of plans and features they offer. Not all hosted SharePoint environments are the same, and certain functions and features available in one plan (such as records management and legal hold) may not be available in another.

[7] Since its introduction in 2001, there have been more than 10 different versions of SharePoint released. As SharePoint's features and functions have evolved and changed, so has the appropriate approach to discovery from SharePoint (for example, the two-stage Recycle Bin was introduced in the Windows SharePoint Services 3.0 and Microsoft Office SharePoint Server 2007 versions of SharePoint.) A helpful discussion of the various current and past versions of SharePoint is provided in Chapter 2 of **Essential SharePoint 2010: Overview, Governance, and Planning** by Jamison, Hanley, and Cardarelli, Addison Wesley, 2011.

[8] For a brief overview of these services see **Essential SharePoint 2010: Overview, Governance, and Planning** by Jamison, Hanley, and Cardarelli, Addison Wesley, 2011, page 35.

(3) Identify the organization's SharePoint experts.[9] Rarely is one individual able to address all questions about SharePoint. It's more likely multiple individuals representing IT (technical) and the business (content) will need to be engaged, as discussed below:

■ **IT (Technical).** Except at the smallest of companies, or where SharePoint is a service to which the organization subscribes, there will be at least one SharePoint technical expert whose title will be something along the lines of System Administrator, SharePoint Administrator, SharePoint Farm Administrator, Web Administrator, SharePoint Specialist, SharePoint Developer, SharePoint Architect, etc. These individuals' responsibilities will vary depending on the size of the company and complexity of the environment, but may include installing SharePoint, updating SharePoint, maintaining the production and non-production environments (servers, storage, and backup.) In some organizations, the technical responsibilities may be separated between application administration and database administration. Technical experts should be able to answer questions about where SharePoint data is stored, how frequently it is backed up, whether there are any system-wide purging or deletion routines, and what (if any) migration of data from legacy systems such as Lotus Notes may have occurred.

■ **Business (Content).** The documents, discussion threads, blog entries, meeting announcements, surveys, and other types of data which reside in SharePoint are the purview of business people. Technical staff may set up and maintain SharePoint, but business people contribute and consume the data it contains. Business people (users) are granted access to SharePoint with varying levels of permissions; permissions determine what the user can do in a SharePoint site. From a discovery perspective, users are critical because they are the ones who compose content and create the data residing in SharePoint—and they are the ones who view and interact with that content and data. There are several categories of SharePoint users, as follows:

† *Site owner.* This person(s) is responsible for assigning rights and permissions to the site's users, and is likely to have substantial knowledge about the site's content. For example, a Team Site for particular project might have as a site owner the project's project manager; this individual would likely have considerable knowledge about the site's content.

† *Members.* These individuals can contribute, consume, and (depending on permissions) delete SharePoint content. This type of user is likely to be found in project and departmental Team Sites where multiple members of an organization are collaborating on a project or functional basis.

† *Visitors.* These individuals can view content but don't typically have the permissions required to contribute or delete content. In a corporate intranet context, visitors would be employees who can access and view content—such as human resource policies—but cannot add to or alter the policies.

---

† *Content contributors.* In certain use cases, such as corporate intranet, there are likely going to be multiple individuals from within the business who compose content to be published via SharePoint. An example would be a staff writer in a company's Corporate Communications department who writes a monthly newsletter about the company's employees (hirings, promotions, retirements, etc.) that is posted on the corporate intranet.

† *Content publisher.* This is an individual who places content into the SharePoint environment. He or she does not typically create content—but rather gathers content and then places it in the appropriate location in SharePoint. The Content Publisher role is typically held by someone with a degree of technical expertise. The term content publisher is not widely used in the SharePoint literature.

† *Hybrid roles.* In some organizations, there may be individuals who bridge the technical, content, and business functions of SharePoint. For example, a Director of Enterprise Content Management might have a strong grasp of SharePoint technology, a clear understanding of SharePoint features and functions, and a direct role determining how SharePoint can improve collaboration and solve business challenges. Also, in some instances, multiple responsibilities are handled by a single individual instead of two or three. For example, a content creator may also be responsible for placing their content into SharePoint.

There may be individuals who bridge the technical, content, and business functions of SharePoint.

---

(4) Understand how SharePoint is configured. This is a daunting challenge, as SharePoint can be used in so many ways, with dozens of different configuration settings which impact where and how SharePoint stores data. Following are key configuration settings of which legal professionals should be aware:

■ **Key general configuration settings.** The following configuration settings are available in both SharePoint Foundation and SharePoint Server 2010.

† *Is self-service site creation enabled?* Self-service site creation is the ability for end-users to create Team Sites on their own; when enabled, it can lead to a proliferation of sites, and with such proliferation technical staff may not be conversant with who has created sites and is storing data in them. Also, those creating the sites may lose interest in the site and forget why they even created it in the first place. Ask custodians and persons of interest to the discovery if they are aware of any potentially relevant SharePoint sites.

† *Is site use confirmation and deletion employed?* If enabled, e-mail messages are sent out periodically to site-owners asking if the site is still being used; if the site-owner does not respond after a certain number of messages are sent (the number is configurable) the site is automatically deleted.

† *Are the Recycle Bins turned on?* SharePoint has a two-stage Recycle Bin, one for each SharePoint user,

---

[9] See **Managing and Implementing Microsoft® SharePoint® 2010 Projects** by Geoff Evelyn, Microsoft Press, 2010, Chapter 5, for a discussion about the various roles and responsibilities for deploying and maintaining SharePoint.

and one for the site collection administrator. They work as follows:

▶ User Recycle Bin. Each user has a Recycle Bin for each site; when the user deletes something from the site, it goes into this Recycle Bin where it remains eligible to be restored by the user for 30 days (number of days is configurable). If a user deletes an item from their Recycle Bin, it is not permanently deleted—the item can be recovered from the site-collection Recycle Bin.

▶ Site-collection Recycle Bin. Site administrators have access to a site-collection Recycle Bin, which contains items deleted by a site's various users. When a user deletes an item, it is placed into two Recycle Bins: the users and the site collections. It is important to note that, when an item is deleted from the site-collection Recycle Bin, it is permanently deleted (although there may be a copy on backup media).

† *Quota.* The quota limits the amount of storage space allocated to each site within a site collection.[10] Once the quota is reached, users may no longer add content to the site. A site quota may cause users to remove content from SharePoint and store it locally—orphaning such content from the SharePoint repository. Such orphaned content, if not identified, represents a risk to the organization.

† *Can SharePoint content be taken offline?* There are several applications which can pull content off of SharePoint, enable the user to view and interact with that content, and then synchronize changes back to SharePoint. SharePoint Workspace, Microsoft Office, and Microsoft Office SharePoint Mobile all offer synchronization with SharePoint, and if used, should be examined to determine if unique content might be resident in these application's storage locations. The ability to take SharePoint content offline can be disabled by the SharePoint system administrator.

■ **ECM configuration settings.** SharePoint's ECM capabilities break down into several sub-categories: document management, records management, digital asset management, and web content management. This article discusses the first two.[11]

† *Document Management.*[12] First, document management functionality is available in both SharePoint Foundation and SharePoint Server 2010. Second, several key configuration settings in document management are:

▶ Versioning—allows the creation of successive versions of a document. From an ediscovery perspective, a question may arise as to whether earlier versions of a document need to be preserved.

▶ Document sets—enables a group of disparate file types to be managed as a single item, with its own metadata. For example, a Word document, Excel spreadsheet, and PowerPoint presentation could be managed as if they were a single file.

† Records Management—[13] records management functionality requires SharePoint Server 2010. Several key configuration settings in records management include:

▶ Information Management Policies

★ *Retention and document expiration.* These settings enable the disposition of content. Policies can be defined which require purging of data upon certain pre-identified events (for example, a contract is purged six years after its expiration date.) Suspension of these settings may be necessary to ensure preservation of potentially responsive data.

★ *Auditing.* Details about when a document was opened, viewed, edited, checked out, moved, etc. can be tracked and later viewed. If enabled, auditing provides information about the activities surrounding content, which could prove pivotal during discovery.

(5) Understand the tools available for the identification, preservation, and collection of data from SharePoint.

■ **Native SharePoint ediscovery tools.**[14] There are two major features built into SharePoint which address ediscovery: search and hold/ediscovery. Search is available in all SharePoint versions (with some noteworthy differences between versions); ediscovery/hold is only available in SharePoint Server 2010.

† *Search:*[15] basic word and phrase searching of site collections is provided in SharePoint Foundation. SharePoint Server 2010 provides more sophisticated search options, including the ability to construct complex Boolean queries.

† *Hold and ediscovery.*[16] This function allows content to be put on legal hold. Content on legal hold cannot be changed or modified in any way.

---

[10] Quotas in SharePoint are similar to quotas in Microsoft Exchange e-mail environments. Where quotas exist, the potential for users to move content from the location with a quota to an alternate storage location exists.

[11] For a complete discussion regarding SharePoint's enterprise content management capabilities, **SharePoint® Server 2010 Enterprise Content Management** by Todd Kitta; Chris Caplinger; Brett Grego; Russ Houberg, published by Wrox, 2011.

[12] For a full discussion about SharePoint's document management features, see Chapter 3: Document Management in **SharePoint® Server 2010 Enterprise Content Management** by Todd Kitta; Chris Caplinger; Brett Grego; Russ Houberg, published by Wrox, 2011.

[13] For a full discussion about SharePoint's records management features, see Chapter 8: Document Management in **SharePoint® Server 2010 Enterprise Content Management** by Todd Kitta; Chris Caplinger; Brett Grego; Russ Houberg, published by Wrox, 2011.

[14] A blog posting which addresses using SharePoint as a platform to support the electronic discovery process is available at: http://ediscovery101.net/tag/sharepoint/

[15] While a full discussion of search in SharePoint is beyond the scope of this article, a more detailed discussion may be found in **Microsoft® SharePoint® Foundation 2010 Inside Out** by Errin O'Connor; Penelope Coventry; Troy Lanphier; Johnathan Lightfoot; Thomas Resing; Michael Doyle, Publisher: Microsoft Press, 2011. See Chapter 2, page 48.

[16] A detailed discussion about hold and ediscovery in SharePoint may be found at http://technet.microsoft.com/en-us/library/ff453933.aspx#How

■ **Non-native SharePoint ediscovery tools.** There are a number of tools promoted as having SharePoint ediscovery capabilities, including:

► FTI (Harvester);

► Nuix (Collector);

► Kazeon (SharePoint Manager);

► AvePoint (DocAve eDiscovery);

► StoredIQ;

► Clearwell; and

► Autonomy (ControlPoint)

Conclusion

The breadth and complexity of SharePoint is intimidating. It's prudent to start as early as possible to gain a reasonable understanding of the target SharePoint environment.

Identifying the version of SharePoint in use, who the various subject matter experts are, gathering information about how SharePoint is used, and how certain features, functions, and settings are configured—these are all things that can be done *now* to help put in place a reasonable plan to execute discovery from SharePoint.

John P. Collins is the Founder and Vice President of Consulting for The Ingersoll Firm LLC. The Ingersoll Firm LLS is the nation's leading ESI data mapping consultancy, working with Fortune 500 companies, AMLAW 100/ 200 law firms, and electronic discovery services providers. Mr. Collins has created dozens of ESI data maps for a wide variety of companies in the pharmaceutical, manufacturing, health services, and consumer products industries. Mr. Collins has an Executive Juris Doctorate from Concord Law School and 16 years of information technology (IT) experience. He can be reached at jcollins@theingersollfirm.com.