

SharePoint and E-Discovery Challenges

By Larry Braggi

Since its launch in 2001, Microsoft's SharePoint application has become many things to many people and organizations. For some, it's a simple collaboration or data management tool. For others, it's a sophisticated development platform for everything from internal dashboards to transactional websites. Numerous case studies available on the Web showcase how companies have used SharePoint to improve worker productivity, strengthen client relationships and help grow businesses.

As the innovative use cases and benefits expand, so does its adoption. In a recently released report, "SharePoint - Strategies and Experiences," the Association for Information and Image Management (AIIM) polled its community of enterprise content management professionals and found that 61 percent of member organizations have implemented, or are in the process of implementing, some version of the SharePoint application (including SharePoint 2003, Windows SharePoint Server and Microsoft Office SharePoint Server).

The same report finds that "the majority of deployments have been initiated without any formal business plan or justification being prepared. The inevitable result is a lack of clarity and planning as to where it will be used, and how it sits with other systems." The report goes on to find that for 23 percent of respondents, all of their office staff access SharePoint, and this is set to double in the next 12 months. In addition, "Forty-four percent of respondents have rolled out SharePoint across 10 or more geographical sites, with 14 percent covering over 100 geographical sites."

These findings are bound to raise red flags across the ECM community for a variety of reasons, but this is especially troubling for organizations that need to collect data in response to e-discovery

requests, whether due to litigation, regulatory requests or internal investigations. Poor planning may be further complicated once SharePoint is deployed and users take advantage of some of the tool's key selling points:

- **Ease of use:** SharePoint was designed to make life easy for the user. Adding, changing and deleting data is fairly easy. It was also designed to let folks be somewhat self-sufficient. The typical installation allows IT to provide a "seed" website to an individual. Once granted, that individual is free to define users, access levels, add content and display that content in a format they like. In addition, they can create child websites below this parent. This makes it extremely difficult for IT staff to monitor and track all these sites, and as a result, a company "guru" for SharePoint content often doesn't exist.

- **Dynamic environment:** Documents and information within SharePoint are changing all the time. Depending on security settings, documents can be added or deleted by the site manager or a common user. SharePoint is designed with collaboration in mind, and it is not unusual for a document to be changed multiple times a day during the drafting phase. Plus, the dynamic nature of the SharePoint environment often reflects the dynamic environment of a company's workforce. As people join a company, access is granted. As people leave a company, access is (hopefully) removed, but there may be a history left behind regarding some of their activity. As people change job responsibilities, they may be granted access to additional content or have access removed (again, potentially leaving history).

Corporate IT teams may already be

working at a disadvantage because they do not know the basic facts about the data in their system, such as where it is, what it is, the total volume, the authors and who had access to it.

Then, for good measure, there are a number of SharePoint configuration features that may complicate the collection process. They include:

- **Date range and author:** In SharePoint, the "Create" and "Last Modified" date and time fields are based on when the document was affected according to SharePoint. So, if Bill created a Word document last week, Sharon copied it to her laptop yesterday, and then Joe added that document to SharePoint today, there would be three different "Create" dates associated with that document. But as far as SharePoint is concerned, the "Create" date and time is today, when Joe added the document, so SharePoint will list Joe as the author. SharePoint is unaware that Bill or Sharon exists.

- **Indexing:** Indexing is a process within SharePoint that traverses the content and stores the words and phrases as a way to facilitate searching. In SharePoint, this can be a global setting like "Turn indexing on for all documents," or it can be controlled at a more macro level, "do not index this website." There are often good reasons for not indexing some sites, such as if content is confidential or the documents are only used for archival purposes. There are "bad" reasons for indexing to be turned off, such as the designer forgot to turn it on. Either way, the problem for e-discovery is that you will not know which sites are not being indexed, because there is no message in the search results.

- **Index updates:** If indexing is turned on, how often does the index get updated with new content and changes? For

example, does the system update the index every night, or every Friday night? This is important to know because content added this morning would not be searchable until indexing has been updated. A good reason for some content not to receive frequent updates is that it simply doesn't change that frequently. An example might be monthly reports that are posted on the first of the month. Once indexed, the documents will not change until next month. But there are some bad reasons indexing is not done frequently, and unfortunately, they are mostly hardware related. For example, if indexing takes too long and interferes with backups or slows down users, it is not uncommon for the time period between indexing updates to get extended. Of course, the more updates are postponed, the more there is to update, just making the matter worse. It is often an easy choice to delay updates to the index when most users don't make use of that feature. From the e-discovery perspective, it is a silent killer.

- **Geographic locations:** According to the AIIM survey, only 15 percent of respondents have a single site. That means 85 percent have SharePoint installed in multiple locations, requiring teams to perform the same search on multiple systems.

Collecting data from any repository will likely run across some common collection issues, such as:

- **Keywords and phrases:** To define keywords and phrases, it is important to understand the issues of the matter. Sometimes defining the issues is easy because there are specific allegations. At other times, the issues cannot be clearly defined until the review is well underway. Changes in the is-

ssues will lead to changes in keywords and phrases; therefore, selecting search terms should be a process that evolves over time. Pick some obvious terms, run searches, analyze the results, adjust search terms and repeat. This is often an iterative process to zero in on the key set of documents, and recent industry trends and case law point to the use of scientific criteria for validating search terms. This is a time-consuming process of testing and documenting results, yet it is necessary to ensure all responsive documents are found.

- **Time delay before preserving:** It usually takes time to define a selection and preservation strategy, including selecting keywords and phrases. During this time, changes are happening to the documents in the SharePoint environment, not only at the document level, but also at the custodian level.

- **Documents unable to be indexed:** Some file types cannot be indexed, or they have little to no searchable content and will not produce search hits as a result. This is to be expected given the nature of many of these file types, such as graphics, sound clips and movie files. However, teams shouldn't be surprised if file types such as PDF or CAD also do not contain searchable text.

While SharePoint is designed to allow searching by keywords, phrases and other criteria, it is a journey best taken with eyes wide open. However, the issue of keyword efficiency is not unique to SharePoint. Legal and IT professionals have worked with keyword shortcomings since it was necessary to put the "e" in e-discovery. Issues are going to change throughout any matter, and it is natural to learn and refine the list of key issues, documents and custodians as things proceed.

Best Practices

The growing adoption of SharePoint makes this a clear target for future e-discovery requests. Yet the dynamic nature of SharePoint, combined with the common delays within the e-discovery process, increases the risk of spoliation. Here are some key tips for legal and IT teams to consider in advance of those inevitable SharePoint collection projects:

- **Research, document and understand your SharePoint environment's configuration.**
- **Make sure your e-discovery policies and procedures extend to SharePoint.**
- **Find a tool that can help map your SharePoint environment, which can provide you with valuable information on the size and types of data included within SharePoint.**
- **Investigate third-party resources that can assist in implementing defensible identification, preservation and collection of SharePoint data.**

Having the above points completed before a matter arises can save valuable time and money, as well as help avoid the risk of spoliation.

Larry Braggi, managing director of FTI Technology, has more than 20 years of experience developing and implementing litigation management and technology solutions for law firms and corporations. In this role he helps clients reduce the cost and risk of e-discovery throughout the entire EDRM process. Larry has been an active leader and participant in user's groups serving in various roles including serving as President of the East Coast Association of Litigation Support Managers (ECALSM) and co-founder of the Etech User's Group.