WILL SHAREPOINT BE THE NEXT GREAT E-DISCOVERY HEADACHE?

Follow-up Q&A provided by Larry Briggi, Managing Director, FTI Technology

CLOUD COMPUTING

Q: What does cloud computing do to the security of docs?

A: In the cloud model, documents are no longer kept on your local drive and often not even on your company servers. Security will be based on the security of the application hosting the data.

Q: In the concept of cloud computing, who owns this data for forensic collection?

A: That depends on the licensing of the "cloud application," but most likely the company paying for the service. In regards to SharePoint, it would be the company. From the point of view of assigning a custodian, you would have to say the Author (who added the document to SharePoint), the person(s) who modified the document, and possibly all the folks who had access to it, even if it can't be proven that they ever opened the document.

CUSTODIANS

Q: What reporting tools are available to report on what a user can access and type actual files being stored by them?

A: FTI offers a Harvester application that traverses the entire SharePoint system and records access at each level (Site, Webs, Folder, List & Document). We also report on file type using the only thing SharePoint knows, which is file extension. For more information on Harvester, please visit www.ftitechnology. com/harvester or email Larry Briggi at harvester@ fticonsulting.com.

Q: Can users themselves set access rights outside of Active Directory? If so, by what -- by user? By document? And how can this be revealed?

A: Yes. Users can potentially configure permissions for all scopes in SharePoint, web application, site collection, site, list, folder or item via direct AD principal (user or group) or SharePoint group assignment within SharePoint.

Q: How does custodian searching work in SharePoint when group access is every bit as transitory as document versions?

A: Because SharePoint is such a dynamic environment, you are limited to collecting "at a point in time." We record information from Active Directory and SharePoint at a point in time and use that information to identify custodian access and authorship. If continuous collections are needed, you would have to augment the initial collection on a "regular" basis, hopefully leaving out documents already collected.

Q: Aside from using the Active Directory to identify custodians who have access to a SharePoint, how do you determine what specific docs are attributed to a particular/specific custodian?

A: SharePoint uses references to Active Directory for its security, so either an individual or group's GUID is listed along with a level of access (read, write, delete…). Authorship is tracked using the user's name.

Q: If multiple people contributed to a document, who's the custodian?

A: In the broadest sense, custodianship is determined by: Author (who added the document to SP), who modified it, and who has access to the document. This is a new area of metadata interpretation, and we hope to see some guidance from the courts. While SharePoint has the ability to track changes (Versioning), this is frequently not turned on. In terms of who modified a document, only the Last Modified Author is tracked without Versioning.

1

Q: How do you recommend collecting data from within SharePoint? Assuming we're doing this by custodian, what procedures do you follow?

A: FTI has created an application specifically to address the identification of documents within SharePoint by custodian. This application copies the selected documents out of SharePoint to a neutral location along with SharePoint metadata and a listing of custodians and their access. As far as I know, this is the only application on the market presenting such a complete solution.

Q: Are you going to talk about how to preserve/ collect data from SharePoint (especially content beyond the simple document libraries and lists)?

A: While SharePoint has been around since 2001, it is still a maturing production in the industry. A wonderful thing about SharePoint is that it is so dynamic it allows designers huge latitude to implement their ideas and create new uses for it. The drawback, at least from the legal side, is that it creates new challenges for identification, preservation, review and production. At FTI, we have taken the approach of responding to the initial needs of our clients in designing our Harvester application (www.ftitechnology.com/harvester). As the industry continues to create new uses of SharePoint, we will address them one at a time. We hope to have another Webinar to discuss some of the issue involved with "non-traditional" data in SharePoint and the handling of duplicates.

Q: When you say "move to a safe environment" to do key word searches, reviewing, etc., do you mean just move to a separate server & view through existing SharePoint tools or do you mean that an electronic discovery tool would be needed?

A: Copying data out of a dynamic environment like SharePoint is an essential step in preservation. A "safe environment" can be a protected network location with very limited access or it can be an external drive which can be taken off-line and locked up. Once this is complete, filtering and culling can be performed on a copy of the preservation data set to narrow the documents moving into the processing and review stage.

Q: Given the growing popularity of SharePoint and the corporate legal and eDiscovery responsibilities to preserve, collect, review, and produce responsive documents, what third party tools would you recommend to enable /automate the eDiscovery processes?

A: Please see the first answer in the Custodian section above.

Q: Concerning vendors, what advice could you give?

A: Ask about their knowledge and ability with handling SharePoint e-discovery throughout the cycle. First, you want someone who can help identify the documents within SharePoint and export copies out (along with metadata and custodian rights). Second, you want someone who can process the data and host it with the SharePoint metadata associated with the review documents. This second part is very tricky as SharePoint is adding what I refer to as a "third layer" of metadata which most processing and hosting tools can't handle. It is important what Web Site a document was part of and who had access. If that metadata can't be accessed in the review environment, your review will be severely hampered.

Q: Selection of a vendor was recommended to identify/collect documents and data. Is it advisable to leave this function in house giving consideration to difficulty of document collection and privilege concerns?

A: Many firms have very bright folks running their networks and SharePoint systems. Unfortunately, they usually don't have experience on the legal side. So even if they can create their own methods for identification and export of data from SharePoint, they will be vulnerable to an outside expert with more experience. E-discovery is the foundation of evidence in a case. If that foundation is shaky, the entire case may crumble. I would submit that until this area of expertise matures, you are more at risk using in-house resources.

SHAREPOINT FEATURES

Q: How does one deal with versioning? Does one produce each version of the document?

A: This is a great question and currently there is no judicial guidance or "correct answer." So, given the context of the case, legal teams should make reasonable conclusions, which can be later defended in court if need be, on which version is most relevant to the matter. In most cases, I would expect only the latest version would be selected for review. With that said, there are good examples of when that wouldn't be the case. For example in a contract dispute, you would want to see all the versions to determine intent. (If the widget was supposed to be blue and the last minute change to green sabotaged the contract.) Or if a matter was limited to HR postings in 2007, you would want the version(s) during that time rather than

the latest. Another side to this question is when a copy of the document exists on two web sites, or exists on a user's C: drive. Can one of these be suppressed as a duplicate? I think we will have to feel out how the industry responds and solutions may need to handle multiple options. We are hoping to discuss some of these topics in a future Webinar.

Q: Is the date-stamping/archiving of documents on SharePoint equivalent to others, e.g. notebooks with dated pages for legal purposes?

A: Creation in SharePoint terms is the date/time a document is added to the system, and it considers the author as the person doing the adding. While this is important, it may not be the document's creation date, nor the original author.

Q: Is it possible for users to set up SharePoint sites that point to content stored on local C: drives? I don't think it is, but want to double check.

A: Yes, it is. Links can point to locations outside of SharePoint on a local drive, on a network share, and even in a third party application like a document management system (DMS). It can also point to documents on someone else's system. Which raises the question, what am I responsible for preserving?

SEARCH/KEYWORDS

Q: What types of documents are NOT searchable in SharePoint?

A: Standard examples of files which would not be searchable are: music, video, & graphic files.

Q: Don't the more sophisticated search tools support faceted or taxonomic searching and alleviate your concerns about search?

A: Concept searching can overcome some of the shortcomings of standard keyword searching, but you still need a starting group of keywords.

Q: If using indexing/key terms to search for relevant docs, how can you best mass export files from SharePoint if you have a large number of files that hit on key terms? Are there good tools for this or does SharePoint have potential for this capability?

A: There are features and tools which will export documents out of SharePoint. The trick is to also obtain and align all the metadata with the exported documents as well as address providing a list of who had access to the documents. In addition, the format of the export

needs to be something which represents the original and is able to be processed, reviewed and produced.

Q: When searching in SharePoint, does the product take into account Parent / Child relationships? ie: Does keyword searching search across attachments?

A: At the simplest level, searching in SharePoint is done by individual document. Unless additional logic is built into the search, SharePoint doesn't think in terms of "families".

Q: What are your recommendations in setting up and maintaining the indexing and search capabilities for SharePoint 2010?

A: That depends on your type of data and your use of SharePoint. As a general rule, you want indexing turned on and updated at "regular" intervals which don't negatively impact daily usage. For example, if data is only added on the first of the month, a single update shortly afterward is all that is needed. For data which is more dynamic, you should adjust and balance against system performance.

Q: Are screenprints, Visio, PDF attachments, etc. searchable within a document?

A: Screen shots are generally graphic in nature and do not contain searchable text. With that said, there are methods of copying text and pasting elsewhere. Visio is searchable, but in my experience, Visio documents in general hold very little text. PDF files come in at least 3 flavors: image only, text only and both. The latter two would be searchable. Adobe provides a SharePoint iFilter tool to assist.

CUSTODIANS

Q: Would you speak to the capturing of communication within SharePoint such as Wikis and Blogs?

A: These types of data are presenting new challenges to the e-discovery industry and for at least the near future, will have to be handled with new creative solutions one at a time. That said, Wiki and blog data contained within SharePoint can be exported, but there are some issues for consideration on how best to format that output for processing and review.

Q: What software does capture the SharePoint metadata?

A: Data in SharePoint is stored in SQL, as such, it can be accessed through the root structure of the SQL database. So, technically, many applications can access the SharePoint metadata. I think the better question is "Which tools understand that metadata and can export it in conjunction with the selected documents?" Unfortunately, most tools don't merge the metadata and documents well for litigation needs. The second question is "How do you handle this extra set of metadata in the downstream processing and review environments?" Again, this is an area where the industry is still evolving. That said, the FTI Harvester tool does export the SharePoint metadata with the documents and our Ringtail Legal hosting environment has been modified to handle this extra metadata so that it is part of the review.

Q: For very large SharePoint sites you may be able to narrow that only a subsite is potentially relevant, but the new SharePoint 2007 tools for backing up (STSadmin) don't handle backing up and restoring some content types in subsites (e.g. surveys). Any tools?

A: In regards to tools for backup and restore, we agree that the existing functionality you mentioned is limited and to our knowledge, there aren't any tools that currently address this. Surveys are especially tricky and may need a custom process which will allow the data to be preserved and presenting in an appropriate context. FTI Harvester does allow selection by site for export and preservation.

SHAREPOINT FEATURES

Q: I assume that if one deletes a document on SharePoint, there will be a record of the date of deletion?

A: Content can remain in the recycle bin, retrievable only by the administrator. Plus, deletion date/time is retained and retrievable if auditing is enabled. However, it is our experience that most sites do not use auditing extensively because of the performance implications.

MISCELLANEOUS

Q: How do you effectuate a solid litigation hold?

A: I would recommend identification of all documents a custodian had access to, authored, or was the last modified author, and exporting a copy of those documents outside of SharePoint. In this way, you have preserved the broadest set of data from a very dynamic environment and reduced your risk of spoliation. If the matter requires an on-going hold, you would augment the initially preserved set with any new documents on a regular basis.

Q: Are there any reported cases dealing with SharePoint discovery directly?

A: As Jim Zucker reported, there are none.

Q: What is the appropriate time period to "augment" collection, presuming you have an ongoing matter?

A: Unfortunately, a blanket rule just isn't possible. An assessment of the appropriate time periods depends on the nature of the matter and the nature of the data which needs to be preserved. While some might say daily is burdensome, in some cases, it may not be frequent enough. In other matters, weekly or monthly may be sufficient.

Q: Wouldn't an entire SharePoint site be subject to discovery since these are typically collaboration areas for particular projects?

A: If an entire site was dedicated to a topic or project and that project was under investigation, you are correct. Preservation with a regular backup may be sufficient, but you still need to export the data and corresponding metadata out for review and production.