Bloomberg Law®

Digital Discovery & e-Evidence®

Best Practices

Not Dead Yet: Custodian-Driven Collection in the Age of Collaborative Spaces

By Farrah Pepper and Andrea L. D'Ambra

What do Mark Twain and custodian-driven collection have in common? In circumstances centuries apart, the rumors of both their deaths had been greatly exaggerated. For custodian-driven collection, the death pronouncements began in recent years, when commentators announced that advances in search technology meant companies could now simply ''search everything'' (presumably by pressing a shiny red button marked ''E-Z Search'') and dispense with custodian-based limitations. The idea behind this premise was that sophisticated litigants had powerful search tools that could index and search the entirety of a company's systems—thereby allowing parties to run search terms across their entire data volume instead of limiting them to a set of agreed custodians. In other words, with the latest nifty tools, the unverified assumption was that one can search all the haystacks in a vast field for that pesky needle.

This is counterintuitive for a number of reasons, not the least of which is the cost of leveraging these tools across data volumes that are increasing exponentially every year. It is far more pragmatic and proportionate to search only the haystacks (or data volumes) that a reasonable investigation indicates may actually contain needles in the first place. Not surprisingly, given the way companies have implemented enterprise search technology, coupled with the new ways information is being created and stored, custodian-driven collection has a new and important role to play in discovery and is far from its deathbed. Rather, in our current data-flooded business world, the emphasis of the amended U.S. Federal Rules of Civil Procedure (''FRCP'') on proportionality requires parties to consider custodian information as *but one filter of several* to be applied to arrive at a proportionate and reasonable collection of potentially relevant information. While the FRCP directly control only U.S. federal litigation, the concepts discussed in this article have broader application to collection and search exercises in a range of legal contexts (e.g., investigations, arbitrations, regulatory responses).

Farrah Pepper is an award-winning attorney with deep expertise in discovery, legal technology, and information governance. Farrah is experienced in building and leading discovery teams, including the GE Discovery COE in her role as GE's first Executive Counsel – Discovery.

Andrea L. D'Ambra is a Partner in the New York office of Norton Rose Fulbright, where she focuses on e-discovery, information risk management, data privacy, and cross-border discovery. Andrea counsels clients on preservation and data management issues and has assisted clients in navigating foreign data privacy laws while complying with U.S. discovery obligations.

How We Work and How We Search

In the last twenty years, the way employees work together has changed significantly. Where once individual team members might have copies only of those documents upon which they personally worked, now companies provide shared network spaces or collaborative applications such as SharePoint, Google Drive, Box, and OpenProject, that allow multiple people in a workplace to access, share and edit the same documents. Although this technology has many benefits in terms of increasing communication and knowledge sharing among employees, it can also create collection

 ISSN 1941-3882

and production issues because it can artificially expand the definition of ''custodial'' data in unhelpful ways that disconnect employees from the data they are actually using. In doing so, this technology potentially increases the amount of data in discovery without increasing the amount of relevant and responsive information.

As a legal matter, companies have a duty to take reasonable steps to identify and preserve relevant information in pending or reasonably anticipated litigation. Because companies act through their employees, generally it is reasonable and proportionate for companies to preserve relevant documents by identifying individual and departmental custodians who have the greatest nexus to the dispute (the ''key players''). Traditionally, once a party identified the custodians likely to have information about the dispute, the party could focus on the data sources where the relevant information would be located: the places the custodians created and stored documents (e.g. custodian email accounts, computer hard drives, and personal network shares). Even as data channels and sources multiplied and became more esoteric—smartphones, tablets, flash drives, departmental network shares—the nexus between the employee and the data source remained strong and the increase in data volumes was generally accompanied by an increase in relevant documents. The workplace craze for collaborative spaces, however, has undermined this nexus.

With a corporate culture emphasizing collaboration and easy access (though data security concerns may be changing this), many company IT departments provide access to shared spaces broadly, sometimes without even notifying the employees. This means the number of data locations that a custodian can access is not only exploding, but the actual connection to the employee is often diminishing because the employee does not avail herself of the access. The ''objective'' data of the IT department shows a broad digital footprint with fingers potentially in many digital pies, but the reality is that the employee has no real connection to many—if not most—of these collaborative data sources. Discovering which documents are relevant and therefore must be preserved is more complicated in workplaces that use collaborative applications, because individual custodians may have access to documents that they have never actually accessed, read, or edited.

The idea of what constitutes a custodian's data continues to evolve. Courts recognize that producing parties have a good faith obligation to search for relevant information and that it is wholly appropriate to rely upon attorneys to execute their ethical duties in this regard. *See Banks v. St. Francis Health Ctr., Inc.*, No. 15-cv-2602, 2015 BL 385487 (D. Kan. Nov. 23, 2015). It makes little sense for requesting parties to dictate where and how a producing party searches, for relevant data for two reasons. First, requesting parties' incentives are misaligned with respect to FRCP 1, which requires ''just, speedy and inexpensive'' discovery, because they tend to want as extensive a search as possible to uncover potentially relevant data. *See Tucker v. Am. Int'l Group, Inc.*, No. 3:09-CV-1499 (CSH), 2012 BL 61214, 281 F.R.D. 85, 95-96 (D. Conn. Mar. 15 2012). In addition, requesting parties generally have little to no knowledge about how an adversary stores and organizes its information. See Fed. R. Civ. P. 26(b)(1) advisory committee's note to 2015 amendment.

As data volumes continue to grow, there is an increasing pressure to use document and workspace metadata to make overarching decisions on the relevance and custody of data sources. This can lead to unintended and inaccurate results, however, when documents are attributed to a custodian who has neither drafted nor read them—but simply had access to the platform or virtual collaboration space where the document was kept. Collaborative spaces are, by their very nature, intended to provide a broad swath of people access to documents related to a project. Some collaboration spaces even allow employees from multiple companies to use a single cloud of data when working on a joint project, further complicating the question of who ''owns'' the data. But access alone cannot be enough to make someone a custodian. Thus, the real question is whether the individual has true *interaction* with the document, not merely access, before determining if the individual may be fairly considered a document custodian.

To put this in perspective, consider the following real world scenario. If a New York employee has an ID card issued by her employer that would allow her to access any of the 100 company offices worldwide, that does not mean that she has ever read, or may be even aware, of the smoking gun budget report sitting in a cardboard box in the Sao Paulo file room. And further, even if the New York based employee swipes into the New York office building every day, she is not considered to ''own'' every scrap of data or information in the building or that may be sitting in a file drawer she passes on the way to get coffee. Simply having access to documents or data would create perverse and disproportionate outcomes if it were deemed sufficient to generate custodial ownership. This is particularly true where the employee does not even avail herself of the limited right granted to read the document for any number of reasons, including discomfort with technology or a preference to work in a more solitary fashion.

Solving the Collaboration Challenge

Having established that collaborative spaces present some distinct challenges to the traditional discovery search and collection paradigm, what, then, are the solutions to ensure proportional and reasonable outcomes when companies need to identify and collect relevant data?

The advent of information governance as a tool to help companies manage their data has offered a partial solution to this challenge. By implementing a comprehensive information governance program, companies can organize and classify their documents to provide helpful programmatic context. But while information governance is an important step in managing a company's information portfolio on a going forward basis, it does not address the challenges around custodial information for existing documents in collaborative spaces unless there is a significant investment in time and resources to classify the existing data. In addition, relatively few companies have truly robust information governance practices. See generally *The Sedona Conference Commentary on Information Governance*, 15 Sedona Conf. J. 125, 131-37 (2014).

One might also wonder if companies could just reconfigure their IT architecture to avoid this problem. In other words, companies should not give access to data in collaborative spaces to employees who are not going

   ISSN 1941-3882

to actively use it, thus eliminating any confusion over whether mere access is sufficient to trigger custodial ownership of a document. But businesses are not in the business of litigation. Although data security considerations may eventually more closely limit access to collaborative spaces, the current zeitgeist favors more unfettered access to encourage sharing, particularly among employees who have not worked collaboratively in the past. It is neither fair nor legally required to expect business and technology decisions that impact employee productivity to be driven by the possibility of legal search requirements that may or may not materialize in the future.

One might hope that greater transparency with a requesting party might lead to positive and rational resolution of these custodial issues around collaborative spaces, but unfortunately, the reality frequently does not meet that ideal. First and foremost, the level of detail a party would need to disclose to engage in a truly helpful discussion is not practical on many levels. Adversaries often use disclosures of some information to demand additional, unnecessary and burdensome discovery based on the unfounded belief (or worse, employing deliberate tactic to suggest) that there may be the chance of relevant information in the additional locations. For example, in *Little Hocking Water Assn., Inc. v. E.I. Du Pont De Nemours & Co.*, the plaintiff demanded that the defense restore archived material to a searchable format and search the ''legacy'' server for documents responsive to its requests. No. 2.09-CV-1081, 2013 BL 80361 (S.D. Ohio Feb. 19, 2013). The court denied the motion and rebuked the plaintiff for seeking discovery that the defendants had confirmed was duplicative of files that were already produced.

Outside demands that a company search additional locations rarely yield relevant data, because the business personnel within the company are in the best position to determine—in the first instance—the most reasonable locations where potentially relevant information might be located. In fact, counsel must invariably speak with the ''key players'' in the litigation in order to understand how they stored information as well as with information technology personnel. *See Zubalake v. UBS Warburg LLC*, No. 02 Civ. 1243 (SAS), 2004 BL 35186, 229 F.R.D. 422, 432 (S.D.N.Y. 2004).

Providing opponents with detail regarding a company's collaborative spaces (whether or not there may be truly relevant data there) spurs opponents to keep asking questions, which in turn stalls the process because the producing party must go back and explain in detail why particular spaces would not reasonably have relevant information. This can also drive up costs, as counsel and business personnel must spend significant time addressing concerns and justifying decisions about data sources where the business already knows there is no potentially relevant information. This potential for a vicious cycle is one of the sound reasons why courts do not allow discovery on discovery absent good cause. See *Mortgage Resolution Servicing, LLC v. JP Morgan Chase Bank, N.A., N.A.*, No. 15CIV0293LTSJCF, 2016 BL 226075 (S.D.N.Y. July 14, 2016).

The complete solution to this challenge is not a simple one, and may require tailoring depending on a number of factors, including the size of the company, the number of custodians from which data must be collected and the configuration of the company's collaborative spaces. Nevertheless, here are some general guidelines to consider in order to ensure proportional and reasonable collection and search in collaborative spaces.

(1) *Metadata has value, but also limitations.* Despite the ascendancy of metadata as a programmatic way to make quick decisions on the relevance and ownership of documents, it has become clear that in certain instances, such as determining custody of a document in a collaborative space, a more thoughtful and investigative approach will yield more accurate results. The techniques developed before reliance on metadata became more widespread, such as custodian interviews, still remain appropriate in instances where a company has good reason to know the metadata of a document is not sufficient.

(2) *Follow the custodial breadcrumbs.* A responding party need look only in those places where potentially relevant information is located—and radiate out from there as further investigation, via custodial interviews and otherwise, points to additional reasonable locations. However, in most cases, and certainly the large ones, interviewing all the custodians will be disproportionately expensive to what is at stake. One might consider sending out a survey form to custodians to solicit from them the locations of relevant documents or, if possible and reasonably practicable, providing the custodian with a list of locations where metadata indicates they have access, so they might isolate only those areas where potentially relevant information would exist.

(3) *Focus on the ''key players'' first.* Even if a party opts to use a survey methodology, a blended hybrid approach may be appropriate, including direct interviews with a few key players up front to gain an understanding of where the company and its custodians generally keep information potentially relevant to the litigation. Fundamentally, the people who created the information generally are in the best position to tell counsel where their documents are located and what information falls within their custody. *See Small v. Univ. Med. Ctr. Of S. Nevada*, No. 2:13-CV-00298-APG, 2014 BL 228669 (D. Nev. Aug. 18, 2014). Once a party interviews and collects for the key players, a proportionate approach would dictate that they need go to collaborative spaces for secondary and tertiary custodians only if and when they determine potentially relevant information is missing and only accessible via that route.

Ultimately, following this reasoned approach when considering true custody of documents in a collaborative space will likely yield more accurate and more proportional results than reliance on metadata alone.

The Way Forward: Targeted Haystacks

As a matter of course in litigation, parties use the identification of custodians whose documents will be searched as a reasonable way to limit the scope of discovery. Given the profusion of data associated with custodians, however, requiring a logical nexus between the custodian and the level of interaction the custodian has with relevant data makes even more sense. It also aligns well with the 2015 amendments to the FRCP that require parties to take into account proportionality considerations when identifying and collecting potentially relevant information. If parties base their collection on solely electronic metadata they know to be inaccurate (for example, relying on overly broad ''access'' to information in a collaborative space to determine if an individual ''owns'' that data), instead of relying upon the

knowledge and direction of the custodians themselves, then the cost of discovery may increase without any corresponding increase (and, in fact, a likely decrease) in the relevance of the information collected and reviewed. In other words, adding more hay to the search does not make finding the needle any easier or more likely.

In sum, custodian-driven collection is still alive and kicking, and with new vigorous application. Rather, it is the knee-jerk reaction to ''search everything'' that is dying, and with good reason. Parties need a reasonable, intelligent, and informed approach to their custodial collections and should not simply by default collect from all locations to which a custodian has access. It is no longer reasonable and proportionate to collect from every place a custodian *could have* stored data; rather, data should be collected from only those places where the custodian affirmatively *did* store data that she interacted with and that is related to the claims and defenses in the matter.

While our friend Mark Twain has long since met the great recycle bin in the sky, we will part with one of his many wise and witty sayings that still have application to the current day. As he opined, ''It's not the size of the dog in the fight, it's the size of the fight in the dog.'' Put another way, nearly anyone can amass a great big pile of hay, so do not measure your discovery success by how big and towering your haystack of data gets. Rather, focus on how many relevant needles you can squeeze out of the proportional and targeted bale of hay your custodians help you find.

***The authors wish to thank Susana Medeiros and CJ Baker for their assistance in researching this article.

  DDEE ISSN 1941-3882