The Computer & Internet Lawyer

Volume 30 ▲ Number 3 ▲ MARCH 2013

Ronald L. Johnston, Arnold & Porter, LLP Editor-in-Chief*

Taming the Fox in the Henhouse: Defensible "Self-Collection" in E-Discovery

By Jeffrey C. Sharer, Colleen M. Kenney, and Sheila A.G. Armbrust

Oceans of ink have been spilled in recent years on the exponential growth of electronically stored information in society and the corresponding impact that growth has had on discovery costs in litigation.[1] All stakeholders in the litigation process—but especially those who find themselves frequently on the responding side of discovery—are looking for ways to harness those costs that are defensible and do not compromise the essential truth-seeking function on which our civil justice system is based. This is no easy task, as the truth-seeking function in the United States is believed best to be served by allowing broad discovery of all matters potentially relevant to the claims and defenses in each case. And in a world where an estimated 89 billion business emails are sent each day,[2] and large organizations are increasingly seeing their data stores break the petabyte barrier,[3]

Jeffrey C. Sharer is a partner in the Chicago office of Sidley Austin LLP, where he concentrates in litigation and enforcement work on behalf of public accounting firms and matters related to electronic discovery, computer forensics, and information governance. Mr. Sharer is a member of Sidley's E-Discovery Task Force; the Sedona Conference Working Group on Electronic Discovery; and the Seventh Circuit Electronic Discovery Pilot Program. Mr. Sharer frequently counsels clients on matters related to electronically stored information, with particular emphasis on the use of best practices and technology to reduce costs and improve both quality of results and defensibility of process throughout the information governance lifecycle.

Colleen M. Kenney is a partner in Sidley's Chicago office and the Chicago Chair of Sidley's E-Discovery Task Force. She is also a frequent speaker and panelist on e-discovery related topics. Ms. Kenney is a Certified Public Accountant, a Certified Management Accountant, and a member of the Sedona Conference Working Group on Electronic Discovery. She has more than 20 years of litigation experience and is a seasoned trial lawyer.

Sheila A.G. Armbrust is an associate in Sidley's San Francisco office. Ms. Armbrust represents clients in a diverse range of criminal and civil litigation and enforcement matters, with particular emphasis on white collar, healthcare, and securities litigation.

Wolters Kluwer
Law & Business

"broad discovery" even in cases with relatively narrow facts and modest stakes can quickly encompass document counts in the tens or hundreds of thousands, or much more.[4]

Not surprisingly, there is no "silver bullet" or "easy button" when it comes to controlling costs associated with electronic discovery. Nor is there any "one-size-fits-all" approach to preserving, collecting, processing, reviewing, or producing electronically stored information (ESI). Rather, the controlling principles of reasonableness and proportionality must be applied to the facts of each case and opportunities to control costs must be evaluated at each stage of the discovery process. Although many aspects of electronic discovery often vary widely across cases, it generally holds true that all else being equal, the larger the volume of data in the e-discovery pipeline at any given stage of a matter, (1) the greater the cost is likely to be at that stage, and (2) the larger the volume (and therefore the greater the cost) is likely to be at later stages in the same matter.[5]

In its seminal 2012 paper "Where The Money Goes: Understanding Litigant Expenditures for Producing Electronic Discovery," the RAND Institute for Civil Justice estimated that in a typical case for the eight large corporations it studied, total e-discovery expenditure generally was around $18,000 *per gigabyte of data reviewed*, with first and third quartiles at $12,000 and $30,000, respectively.[6] Total e-discovery costs overall on the 45 cases for which data was available ranged from a modest $17,000 on one end to a jaw-dropping $27 million on the other, with a median value of $1.8 million.[7] The Institute further found that the major cost driver—estimated to account for at least 70 percent of expenditures—was review for relevance, responsiveness, and privilege, and concluded that "review costs would have to be reduced by about three-quarters in order to make those costs comparable to processing, the next most costly component of production."[8] It continued: "Choosing a 75-percent reduction in review expenditures as the desired target is an admittedly arbitrary decision, but more-modest cost savings are not likely to end criticisms from some quarters that the advent of e-discovery has caused an unacceptable increase in the costs of resolving large-scale disputes."[9]

The Institute explored at length possible ways of reducing review costs.[10] Among other things, it observed that labor costs associated with review "may well have bottomed out, with further reductions of any significant size unlikely," and further that given the tradeoff between review speed and comprehension, "it is unrealistic to expect much room for improvement in the rates [*i.e.*, documents per hour] of unassisted human review."[11] And although the Institute concluded that "predictive coding" and other computer-assisted review technologies have the potential to identify "at least as many documents of interest as traditional eyes-on review with about the same level of inconsistency," and possibly to do better, at costs that "are likely to be substantially lower than the costs of human review,"[12] such technologies are not inexpensive and often are priced per unit (*i.e.*, per document or per gigabyte) processed.[13]

Thus, it remains that the surest way for any party not to incur the cost—at any level—of reviewing irrelevant data is not to collect it in the first instance. Put another way, by increasing precision at the collection stage, the party avoids not only the cost of collecting irrelevant data but also the costs of processing, reviewing, and potentially producing such data. The challenge, of course, is to increase precision without sacrificing recall; that is, in any given collection, to reduce the volume of irrelevant data retrieved without also reducing the volume of relevant data retrieved.

Enter "self-collection." Although the term does not have a single, universal definition, generally it denotes that a party is relying in some manner on individual custodians (*e.g.*, a corporation relying on its employees) either to identify and/or to copy or otherwise provide to counsel those documents among their files that are potentially relevant to the litigation. Nothing about "self-collection" is unique to ESI; to the contrary, litigants and lawyers have relied on individual custodians for document collection since the days when the few computers that existed were the size of city blocks, and in the so-called paper world, no one seriously would have proposed photocopying an entire warehouse so that it could be searched later for a small subset of potentially relevant documents.[14] What is different about "self-collection" today is that the scale of discovery is such that relevant documents are no longer subsets of warehouses, they are subsets of what amounts, in digital form, to tens or hundreds or thousands of warehouses. As a result, "the collection process has necessarily had to adapt to the rapid changes and volume considerations involved," giving rise to, among other things, a need in almost all cases "to engage IT and business professionals who are knowledgeable about the sources and locations of ESI within the enterprise."[15]

When used properly and under the right circumstances (two significant conditions discussed further below), custodian self-collections can be appropriate, efficient, and cost-effective, allowing parties, among other things, quickly to identify, review, and produce documents that are relevant to the matter and

to avoid wasting time and money on documents that are not relevant. However, a survey of case law and commentary in recent years could easily give the impression that the term "self-collection" is a four-letter word in the e-discovery context.[16] Several courts have imposed sanctions—some severe, and appropriately so—on litigants for discovery lapses involving self-collection that has been inappropriate, poorly executed, or both.[17] Those cases have been noticed. With headlines such as "Self Collections in E-Discovery–Just Too Risky for Prime Time,"[18] "How Dangerous Is Self-Collection in E-Discovery?,"[19] and "Judge Scheindlin Says 'No' to Self-Collection,"[20] many commentators have reported strong judicial disapproval of "self-collection" and have concluded along similar lines that "th[e] approach is simply far too dangerous for most enterprises, except perhaps those that are extremely risk tolerant."[21]

It is a belief in many corners that employees are sufficiently likely to be biased, or to conceal information that is personally damaging or embarrassing.

The rationales most often cited by courts and commentators critical of "self-collection" are twofold. First is a belief in many corners that employees are sufficiently likely to be biased, or to conceal information that is personally damaging or embarrassing, that relying on them to identify potentially relevant documents is akin to having the metaphorical "fox guarding the henhouse."[22] Second is that without proper direction and supervision, an average employee has neither the legal nor the technical expertise needed to identify and/or acquire potentially relevant ESI for purposes of litigation.[23] Although both concerns have merit and, in certain cases, can limit or preclude the sort of reliance on custodians that has been labeled "self-collection," the risk of discouraging such reliance in all cases is that parties will be led to collect ESI more broadly than reasonableness and proportionality require, for example, imaging that captures entire hard drives or entire email stores when more targeted collections would suffice. Then the parties would face significantly greater costs to move that oversized data set through the processing and review phases of the discovery process. Beyond the direct impact that over-collection has on costs in the immediate matter, are the indirect but no less substantial costs that many organizations incur over the long term as a result of having to preserve, and potentially search and review, over-collected data that otherwise would have expired in the normal course of business, in connection with future matters.

A closer reading of the case law suggests that in many cases, neither the fox-and-henhouse concern nor the other failures that have led to sanctions associated with "self-collection" is intractable, insurmountable, or even daunting. In many cases, targeted collections that rely heavily, or even entirely, on custodians to "self-identify" potentially relevant documents among the ESI within their personal worksphere can be both reasonable and defensible.[24] Even protocols that call on custodians to segregate, copy, or otherwise "self-acquire" or "self-harvest" their own potentially relevant data so that it can be advanced to the processing and review phases of the discovery process likely have a place in the right types of cases. Furthermore, although there certainly are circumstances in which the fox-and-henhouse metaphor is apt, in practice such cases are more the exception than the rule and, moreover, usually will be readily apparent to thoughtful parties and counsel. As a result, any party faced with collecting ESI from a number of custodians, before reflexively undertaking costly and overbroad processes such as imaging entire hard drives or servers, should consider whether a targeted approach carefully designed, properly supervised, and well documented that relies heavily or even entirely on the inherent familiarity that each custodian has with his or her data is appropriate.

What's in a Name? Let's Be Clear

Some of the uncertainty regarding the appropriateness of "self-collection" is a result of the term itself being used loosely to mean multiple things. As noted above, "self-collection" broadly encompasses reliance on individual custodians to support the collection process in one or both of two related-but-different ways: (1) to "self-identify" potentially relevant documents within the custodian's personal worksphere, *e.g.*, in the custodian's email, stored on the custodian's local hard drive or removable media, or in the custodian's assigned network storage; and (2) to "self-acquire" or "self-harvest" such documents by segregating, copying, or otherwise capturing them to be advanced through subsequent stages of the discovery process. In either respect, the nature and degree of reliance placed on the individual custodian can vary widely. For example:

- At one end of the continuum, a party forwards a complaint and discovery requests to its employees with cursory instructions to read the requests, locate potentially relevant documents, and forward copies to counsel.

- Toward the opposite end of the continuum, a party works with counsel to distill and summarize the relevant pleadings and discovery requests in terms that can be understood clearly by non-lawyers; distributes that guidance within the organization; interviews key custodians to confirm that they understand what's required; and then provides step-by-step instructions to walk each custodian through the process of copying the potentially relevant data in a manner that preserves, to the extent necessary in the particular case, the data's forensic integrity.

In both cases, the organization is relying heavily on the custodian to "self-identify" potentially relevant documents. Likewise in both cases, the organization asks the custodian to "self-acquire" or "self-harvest" the documents that the custodian identifies. In both respects, the approaches can be described in whole or in part as "self-collection." However, there is no question that the two approaches differ significantly and in important ways, and that all else being equal, the second is more likely than the first to withstand scrutiny in the event that relevant documents are missed.

The distinction between relying on custodians to ***self-identify*** potentially relevant documents, on the one hand, and relying on custodians to ***self-acquire*** or ***self-harvest*** such data for purposes of advancing it to the processing and review phases of the discovery process, on the other, is significant. Identification usually entails communicating the substance of a discovery request to the custodian and asking the custodian to determine whether he or she has any data that falls within the request and, if so, where that data is stored—all topics about which most custodians in most cases will be especially, if not uniquely, knowledgeable.

Acquisition, or harvesting, entails copying identified data for purposes of further processing, review, and potentially production. Unlike identification, where the custodian usually possesses at least some informational advantage over the organization and counsel, acquisition fundamentally involves IT and/or digital forensics, areas in which, if anything, the organization, counsel, and/or an e-discovery vendor usually will know more than the custodian. This is not to say that self-acquisition, or self-harvesting, cannot be a reasonable and proportional approach to collecting ESI from certain types of custodians or in certain types of cases; to the contrary, it can be. Rather, it is simply to underscore that although either or both self-identification, on the one hand, and self-acquisition or self-harvesting, on the other, may be referred to as "self-collection,"[25] each presents its own legal, technical, and practical issues and therefore must be considered independently of the other.

Foxes and Henhouses, Babies and Bathwater, and the Importance of Process

As noted above, the legal standards by which a party's discovery responses, including collection, are measured are easy to state: reasonableness and proportionality.[26] Furthermore, Sedona Principle No. 6 recognizes that "[r]esponding parties are best situated to evaluate the procedures, methodologies, and technologies appropriate for preserving and producing their own electronically stored information."[27] Determining whether the "procedures, methodologies, and technologies" employed by a particular responding party in any given case were reasonable and proportional, however, is not a simple task; both reasonableness and proportionality are intensely fact-specific inquiries involving considerations related to the litigation and to the party's human, IT, and other resources for which it is difficult, if not impossible, to discern blanket answers from case law. That said, a careful reading of those decisions that have addressed various forms of "self-collection" reveals at least some common themes that, in turn, can be used to inform whether and when it is appropriate for an organization to rely on its employees in connection with the collection process and, where some degree of reliance is appropriate, what steps the organization can take to maximize the likelihood that its processes will be found reasonable and defensible in the face of a later challenge.

Self-Collection Should Not Be Used When There Is in Fact a "Fox in the Henhouse"

One message that has been sent loudly, clearly, and consistently by courts is that when an employee has a personal stake in the dispute, such as when the employee is alleged to have committed the bad acts on which liability is asserted, it is unreasonable for a party to rely on that employee either to self-identify or to self-acquire potentially relevant documents. This is the circumstance that invokes, aptly, the fox-in-henhouse metaphor. When parties have relied on interested employees to self-collect potentially relevant documents, courts have effectively held those parties to strict liability for errors and omissions by consistently deeming reliance under such circumstances unreasonable, or worse.

In *Suntrust Mortgage, Inc. v. AIG United Guaranty Corporation*,[28] the Eastern District of Virginia sanctioned a plaintiff, awarding the defendant attorney fees and costs, based on discovery abuses arising from an interested employee's alteration of emails germane to

the issues disputed by the parties. Specifically, Suntrust had entrusted the employee, whose individual conduct was a central issue in the litigation, to self-collect emails related to her alleged wrongful conduct.[29] Even after senior officers and in-house counsel determined that there were important differences in content between emails produced by the interested employee and the same emails provided by defendants, they waited several months to interview the interested employee about the discrepancies.[30] The court also observed that, although Suntrust "imaged the hard drive of [the employee's] work computer … [Suntrust] did not hire forensics experts or any other kind of outside help to assist in the collection and analysis of [the employee's] electronic files."[31]

When an employee has a personal stake in the dispute, it is unreasonable for a party to rely on that employee.

Other courts have sanctioned parties for lapses attributable to self-collection by interested employees even absent evidence or allegation of deliberate misconduct. For example, in *Jones v. Bremen High School District 228*, an employment discrimination action, the Northern District of Illinois held that the defendant school district was reckless and grossly negligent when it allowed three employees—one of whom "was at the center of plaintiff's complaints"—to search their own email "without help from counsel and to cull from that email what would be relevant documents."[32] The court explained:

> It is unreasonable to allow a party's *interested employees* to make the decision about the relevance of … documents, especially when those same employees have the ability to permanently delete unfavorable email from a party's system. … Most non-lawyer employees, whether marketing consultants or high school deans, do not have enough knowledge of the applicable law to correctly recognize which documents are relevant to a lawsuit and which are not. Furthermore, employees are often reluctant to reveal their mistakes or misdeeds.[33]

Accordingly, the court imposed sanctions, including a monetary award, jury instruction, and preclusion of certain arguments at trial.[34]

A clear lesson to be taken from *Suntrust*, *Jones*, and similar cases is that when an employee is alleged to have been involved personally in misconduct, or the organization has some other reason affirmatively to question the employee's objectivity or trustworthiness, reliance on the employee to self-collect is almost certain to be found unreasonable in the event relevant documents are missed. But in the absence of any indication that the employee has a personal stake in the controversy, or is otherwise biased or untrustworthy, the fox-in-henhouse metaphor is inapt and, as such, should not foreclose or even constrain reasonable reliance, properly planned and executed, on the custodian's inherent familiarity with his or her electronically stored information.[35] To answer one metaphor with another, in the absence of some indicia of unreliability, don't throw the baby out with the bathwater.

Self-Collection Requires Clear Direction and Supervision by Counsel and Should Be Carefully Documented

Other courts have sanctioned parties not because they permitted custodial self-collection, but because the collections were not properly directed, supervised, or documented. For example, in a decision widely noted for the unusual, and unusually harsh, sanction imposed, the Eastern District of Texas in *Green v. Blitz U.S.A., Inc.*,[36] a products liability action, imposed civil contempt sanctions of $250,000 and ordered defendants to file a copy of the sanctions order in every lawsuit in which it had been involved during the two years prior and would be involved in the five years subsequent. It did so after plaintiff's counsel received, through discovery in another lawsuit involving a different client suing on the same product, "numerous documents that … [were] extremely relevant and material" and that had not been produced to plaintiff in the case at bar.[37] Through ensuing evidentiary hearings, the court found that the defendant had designated one employee solely responsible for collecting and producing relevant documents for discovery, and that employee testified: "I am about as computer literate—illiterate as they get."[38] Making matters worse, the same computer-illiterate employee, on whom the defendant had relied almost exclusively to find and produce relevant documents over approximately a four-year period, had been personally responsible for the defendant's research and investigation surrounding the alleged product defect on which the plaintiff's claim was based.[39] Yet despite the employee's direct responsibilities related to the alleged product defect—among other things, he was one of three recipients of a key email that the court found "[a]ny competent electronic discovery effort would have located"—and self-professed lack of technical ability, there was no evidence that the defendant or counsel did *anything* to guide or supervise the employee's collection efforts.[40] In the court's view, had the defendant or counsel supervised the employee's work with

"[a]ny competent … effort," it would have identified the problems upon which sanctions ultimately were based and its failure to do so was willful misconduct that warranted the harsh sanctions imposed.[41]

Judge Scheindlin's recent opinion in *National Day Laborer Organizing Network v. United States Immigration & Customs Enforcement Agency*,[42] likewise underscores the importance of supervision by counsel and careful documentation of collection efforts. In *National Day Laborer*, five governmental agencies were served with Freedom of Information Act (FOIA) requests to produce documents in connection with 2008 immigration policies.[43] The parties cross-moved for summary judgment regarding the adequacy of the agencies' searches, and plaintiffs argued in opposition to the agencies' motions that the agencies' searches were inadequate and had not been appropriately documented.[44] Judge Scheindlin denied in part the motions of three federal agencies on the grounds that these agencies either had conducted "woefully inadequate" searches or had documentation "insufficient to permit proper evaluation."[45] In language that has been widely quoted by commentators, Judge Scheindlin answered the agencies' question "why custodians could not be trusted to run effective searches of their own files, a skill that most office workers employ on a daily basis," by stating that:

> First, custodians cannot "be trusted to run effective searches," without providing a detailed description of those searches. …The second answer to defendants' question has emerged from scholarship and caselaw only in recent years: most custodians cannot be "trusted" to run effective searches because designing legally sufficient electronic searches in the discovery or FOIA contexts is not part of their daily responsibilities.[46]

Despite its result and characteristically strong words from Judge Scheindlin, *National Day Laborer* is not a death knell for self-collection. Far from it. *National Day Laborer* stands primarily for two unremarkable propositions. First, because search and retrieval *for purposes of litigation* is not part of an average employee's daily responsibilities, it is essential that sufficient direction and supervision be provided to the employee by the organization and counsel.[47] Second, it is essential that whatever steps a party takes to discharge its e-discovery obligations, they should be carefully documented so that they can be proved to a court's satisfaction in the event of a later challenge.[48] Indeed, Judge Scheindlin *granted* summary judgment to two federal agencies that also had relied primarily on custodial self-collection in their productions, where one agency had provided "specific mandatory search terms to custodians and confirmed to the Court that the custodians used those terms," and the other had "provided the precise terms that its employees used to search individual and shared sources … ."[49]

Best Practices for Custodian-Assisted Collection

Once a party has determined that it is appropriate in a given case to rely on individual custodians to self-identify and/or to self-acquire potentially responsive ESI, it must then determine how to go about this. Although this always is a fact-specific determination based on the organization's needs, systems, resources, and other factors, it is nevertheless possible, based on the case law and other guidance discussed above, to discern some best practices that can be applied in most cases and that, if applied, can stave off discovery challenges or enhance the defensibility of the collection process in the event of a challenge.

First, as with most activities involving electronic discovery, a party contemplating self-collection should consider whether cooperation and transparency are appropriate.[50] This could involve, among other things, mutual identification of custodians, search terms or protocols, or tools or processes to be used for acquisition. If a relevant document is missed but the producing party had, or sought reasonably and in good faith, the requesting party's agreement to the protocols used, it will be difficult for the requesting party to complain in the event the document later surfaces.

Second, it is essential to any form of self-collection that the custodian be provided with clear and concise direction, in form and substance that can be readily understood and applied by a non-lawyer, regarding what the custodian is being asked to do.

- If the custodian is being asked to self-identify potentially relevant documents, it is important that the custodian understand what "potentially relevant" means; although most employees probably can identify, as potentially relevant to a lawsuit, documents that on their faces clearly reference the subject matter of the dispute, as courts have admonished, the scope of civil discovery is much broader than the obviously relevant, and if the organization is going to rely on the employee's identification, it must ensure that the employee understands the full scope of what needs to be identified.[51]

- If the custodian is being asked to copy or otherwise extract ESI from his or her computer or

other devices to be forwarded to counsel (such as by "drag and drop" copying to external media or "exporting" data from its native application), or to do anything else that is likely to affect the contents of the ESI or its associated metadata (such as by moving potentially responsive files from their current locations to a dedicated "collection" folder on local or network storage), it is important that the custodian either have or be provided with the technical understanding needed to maintain, insofar as is needed in the particular case, the forensic integrity of the ESI.[52]

Third, it also is imperative that counsel manage and supervise the collection process to ensure that appropriate guidance and instruction is provided; that chain-of-custody is properly followed and documented; and that the results of the process are reliable and defensible. In most cases, this will involve custodian interviews conducted by counsel, at least for "key players," along with individualized follow-up as needed and reasonable and proportionality quality control processes.[53]

It is essential to any form of self-collection that the custodian be provided with clear and concise direction.

Fourth, a party relying on custodian identification and/or acquisition of potentially relevant data should consider whether the custodian's efforts should be supplemented by secondary processes, essentially resulting in a "hybrid" approach where primary reliance is placed on the custodian to self-identify and possibly self-acquire the documents, but the secondary process serves as a "safety net" or "backstop." For example, when using forensic software to collect files and folders that have been identified by a custodian from the custodian's hard drive, it may be possible and, in some cases, prudent to have the software also index and apply targeted search terms to the balance of the drive, *i.e.*, to the data not identified by the custodian, to capture items potentially overlooked by the custodian. Priority still can be given in processing and review to the custodian-identified data, with the keyword-identified data serving only as a safety net and/or to validate the completeness of the custodian's designations.[54] A similar result can be accomplished through the use of centralized collection tools, which the RAND Institute observed are "quite powerful, providing an automated collection process across the company's internal network without directly interrupting work being performed by a targeted custodian or data location, but [ ] require a fairly significant upfront investment of money and labor … and are only now becoming standard in large companies."[55]

Fifth, as discussed above, there may be cases in which it is appropriate to have individual custodians copy or otherwise transfer their potentially responsive ESI to external media. However, in most large cases there is an expectation, and in many cases a requirement, that responsive ESI be produced in a form that maintains in full the forensic integrity of the original file, including modified-accessed-created dates (commonly referred to as MAC or, with entry-modified, MACE dates) and other system metadata, some or all of which may be affected by most of the processes by which an average custodian can duplicate ESI. This means that in most cases, even if it is entirely appropriate to rely on individual custodians to identify potentially relevant data on their computer systems and devices, it may still be necessary either:

- To have a forensic specialist (which can be a third-party vendor or, if the party has the capability in-house, an internal resource) copy or otherwise acquire the data using accepted forensic tools; or

- To provide to custodians the tools, as well as the technical direction, support, and supervision, necessary for custodians themselves to create forensically sound copies of targeted files; for example, so-called "plug and play" self-collection tools are available from many established companies including Guidance Software (developer of EnCase); AccessData (developer of Forensic Tool Kit, or FTK); IKON Litigation Support Solutions; and others.[56]

Sixth, just as all cases are not created equal, neither are all custodians. Accordingly, the party and its counsel should consider the appropriateness of custodian involvement in the discovery process separately for each custodian or category of custodians in a given case. It may be that there is a core group of custodians for whom it is necessary, for example, because of their importance to the dispute or the nature of their involvement, to collect broadly, for example, imaging their entire hard drives, but that some form of self-collection is appropriate for the rest of the population. Counsel also may determine that certain individuals outside the company's control, namely former employees or contractors, may not be appropriate candidates for self-collection. By taking a tiered approach, the organization can achieve the significant cost and other benefits associated with self-collection for many of its key custodians while

simultaneously "playing it safe" with respect to those custodians for whom self-collection would involve heightened or undue risk.

Seventh, counsel should consider preserving broadly despite collecting narrowly, particularly at the outset of litigation. Thus, counsel should ensure that it issues a sufficiently broad legal hold, and gives due consideration to suspending, at least temporarily, the rotation or recycling of any media that is likely to contain unique, potentially relevant information, until a time when counsel has provided sufficient guidance to custodians who will self-collect and counsel has comfort with the effectiveness of self-collection.[57]

Eighth, counsel must timely audit self-collections and adjust course as necessary and appropriate. When custodians begin the process of self-collecting, custodians or counsel may identify a need to refine or supplement guidance. Counsel also may need to ask a custodian to collect from additional data stores or apply additional search terms, as warranted by the litigation. Counsel may even change course on permitting a particular custodian to self-collect, based on challenges that the custodian may face in understanding the substance or procedures inherent in the task of self-collecting.

Ninth, as with most activities involving electronic discovery, the importance of contemporaneous documentation cannot be overstated.[58] The nature of litigation is such that issues that may be perceived by a requesting party to be defects in the producing party's processes may not be identified until long after the underlying collection, processing, review, and/or production has been completed. For example, witnesses routinely are asked in depositions what steps they took to preserve and/or collect documents potentially relevant to the pending litigation, and answers to such questions often lead to follow-up inquiries and, in some cases, disputes regarding the adequacy of what was done. In many cases, however, depositions often do not take place until after document discovery has been completed, and, especially in large cases, document discovery can stretch over periods ranging from a few months to several years. As a result, many disputes regarding electronic discovery involve actions taken, or not taken, many months or even years earlier, and when such disputes arise, the existence of documentation prepared contemporaneously with the events at issue, in addition to being highly probative evidence in its own right, can be the difference between a crucial affidavit or testimony being possible or not.[59]

Accordingly, as it relates to self-identification and/or self-acquisition of potentially responsive documents, the organization and counsel should consider documenting, with respect to each custodian or class of custodians being relied on, the following:

a. Identity of custodian or class of custodians;

b. Nature of involvement with matters at issue in litigation;

c. Whether self-identification, self-acquisition, or both is appropriate and rationale for determination;

d. Documents or categories of documents sought from custodian;

e. Written instructions and other guidance and supervision provided by counsel to custodian, including but not necessarily limited to documents sought, locations to be searched, protocols or methodologies for conducting search, and protocol for designating and/or copying documents identified as potentially responsive;

f. Results provided by custodian to organization or counsel; and

g. Quality assurance procedures and results with respect to reliance placed on custodian.

Conclusion

The digital universe is growing exponentially and shows no signs of slowing. Even with machines able to categorize data more quickly than humans, the costs associated with collecting, processing, reviewing, and producing documents in litigation are likely to remain a source of considerable pain for litigants (and especially for serial litigants) into the indefinite future. The only way to reduce that pain to its minimum is to use all tools available in all appropriate circumstances within the bounds of reasonableness and proportionality to control the volumes of data that enter the discovery pipeline.

Custodial self-collection, either in the form of self-identification and/or in the form of self-acquisition or self-harvesting, can be a useful tool on a responding party's pegboard. The case law makes clear, however, that it is a tool that must be used thoughtfully and in appropriate cases. Among other things, this means being alert to circumstances in which custodians have personal interests or other indicia—in fact or appearance—of bias or untrustworthiness, and taking care to provide the level of direction and supervision needed to make reliance on custodians' efforts reasonable. But as long as due care is taken in planning, executing, and documenting all aspects of the collection process, parties often will be able to tame the fox in the henhouse and make "self-collection,"

especially but not only in the form of self-identification, a significant and potentially valuable component of a reasonable and proportional electronic discovery plan.

Notes

1. *See*, *e.g.*, Jason R. Baron, "Law in the Age of Exabytes: Some Further Thoughts on 'Information Inflation' and Current Issues in E-Discovery Search," XVII *Rich J. L. & Tech* 9 (2011); *see also* John Gantz & David Reinsel, "The 2011 Digital Universe Study: Extracting Value From Chaos," at 1-2 (IDC iView Jun. 2011), available at *http://idcdocserv.com/1142* (estimating that in 2011, the amount of digital information created and replicated would surpass 1.8 zettabytes, or 1.8 trillion gigabytes, having grown by a factor of nine in five years; also noting that although 75 percent of information in digital universe is generated by individuals, enterprises have some liability for 80 percent of information in digital universe at some point in its digital life).
2. The Radicati Group, Email Statistics Report, 2012–2016—Executive Summary, at 2-3 (Apr. 2012), available at *http://www.radicati.com/wp/wp-content/uploads/2012/04/Email-Statistics-Report-2012-2016-Executive-Summary.pdf.*
3. A petabyte is 1,000 terabytes or 1,000,000 gigabytes, and has been equated to approximately 20 million four-drawer file cabinets filled with text or 13+ years of HD-TV video; the entire written works of mankind, from the beginning of recorded history, in all languages, is estimated to be 50 PB. *See* The Mozy Blog, "How Much Is A Petabyte," available at *http://mozy.com/blog/misc/how-much-is-a-petabyte/* (Jul. 2, 2009); *see also* Joseph McKendrick, Big Data Is Real and It Is Here: 2012 Survey on Managing Big and Unstructured Data at 4-5 & Fig. 1, *MarkLogic* (Apr. 2012) (reporting that as of January 2012, "12 percent of respondents report that they support more than a petabyte of data, and another 32 percent say they have data volumes in the hundreds of terabytes").
4. Consider that even a "budget-friendly" business laptop today usually ships with a hard drive sized at 250 GB, 320 GB, or more (*see*, *e.g.*, Dell Latitude E5430, available at *http://www.dell.com/us/enterprise/p/latitude-e5430/fs)*, with each GB capable of storing roughly a pickup truck full of paper (*see*, *e.g.*, E-Discovery Team, "How Much Data Do You Have?," available at *www.e-discoveryteam.com* (sidebar)). And although empirical data regarding volumes of data collected in e-discovery is scarce, a 2010 survey of law firms, corporations, consultants, and software providers reported, for laptop and desktop hard drives, a mean of 22 GB and a median of 18 GB collected per source. *See* Dutton LLC, "eOPS 2010: Electronic Discovery Operational Parameters Survey," available at *http://www.catalystsecure.com/blog/wp-content/uploads/2011/07/Electronic-Discovery-Operational-Parameters-Survey.pdf* (Apr. 2010). Using a conservative estimate of 2,500 documents per GB, translates to approximately 55,000 documents collected per custodian hard drive; multiply that number by even a modest number of custodians and the counts can quickly become staggering. *Seeid.* at 5 (mean reported average documents per GB is 5,244 and median is 5,500); *see also* John Tredennick, "Shedding Light on an E-Discovery Mystery: How Many Documents In a Gigabyte?," available at *http://www.catalystsecure.com/blog/2011/07/answering-an-e-discovery-mystery-how-many-documents-in-a-gigabyte* (Jul. 7, 2011) (analyzing 5.5 TB of data from 29 cases and estimating approximately 2,500 documents per GB); EDRM, "EDRM Evergreen/Processing/Analysis and Validation," available at *http://edrm.net/wiki2/index.php/EDRM_Evergreen/Processing/Analysis_and_Validation#endnote_images* (Jan. 31, 2008) (reporting results of "industry benchmark survey" with low, median, and high counts, respectively, per GB at approximately 9,900, 22,500, and 36,500 for email, and at approximately 7,500, 15,800, and 20,300 for application files).
5. *See*, *e.g.*, RAND Institute for Civil Justice, Where the Money Goes: Understanding Litigant Expenditures for Producing Electronic Discovery, at 26 & Fig. 2.8 (2012) ("Generally, as collection size increases, so does the final total for production.").
6. *Id*. at 19-20 & Fig. 2.2.
7. *Id*. at 17-19 & Fig. 2.1.
8. *Id*. at 25-27 & Fig. 2.6.
9. *Id*. at xvi.
10. *See id*. at 41-84 (Chapters 3 and 4).
11. *Id*. at xvi; *see also id*. at 43-52.
12. *Id*. at xviii; *see also* Maura R. Grossman & Gordon V. Cormack, "Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review," XVII *Rich. J.L. & Tech*. 11 (2011); Herbert L. Roitblat, Anne Kershaw, & Patrick Oot, "Document Categorization in Legal Electronic Discovery: Computer Classification vs. Manual Review," 61 *J. Am. Soc'y Info. Sci. & Tech*. 1 (2010).
13. Jonathan Berman, "A Dispatch from the Front Lines of E-Discovery," available at *http://www.law360.com* (Oct. 10, 2012).
14. As The Sedona Conference explained in 2009:

    > Until recently, lawyers well knew how to ask for and collect "documents." Key custodians would be asked to gather their hard copy documents and files into boxes, which were made available to lawyers or paralegals to review essentially each and every page for relevance and privilege reviews. This time-worn process admittedly grew more complex in large litigations, *e.g.*, antitrust actions or products liability class actions, where tens or hundreds of thousands of boxes of documents were collected from a corporate enterprise, to be reviewed by legions of junior and contract attorneys. Much the same process continues to be employed today for reviewing huge bodies of evidence that exist only in hard-copy form.

    The Sedona Conference, "Commentary on Achieving Quality in the E-Discovery Process," 14 (May 2009 Public Cmt Ver).
15. *Id*.
16. *See, e.g.*, Sheila McKay, "How Dangerous Is Self-Collection In E-Discovery," available at *http://ediscoverytalk.blogs.xerox.com/* (Jul. 30, 2012); James D. Shook, "Self-Collection Is Dead (Long Live Self Collection!)," available at *http://www.kazeon.com/blog/* (Jul. 2012); Ralph Losey, "Another 'Fox Guarding the Hen House' Case Shows the Dangers of Self-Collection," available at *http://e-discoveryteam.com/* (Mar. 2011).

17. *See*, *e.g.*, Green v. Blitz U.S.A., Inc., 2011 WL 806011, 2:07-CV-37 at *10 (E.D. Tex. Mar. 1, 2011); Jones v. Bremen High School Dist 228, 2010 WL 2106640, No. 08 C 3548 at **9-10 (N.D. Ill. May 25, 2010).

18. Dean Gonsowski, "Self-Collections in E-Discovery—Just Too Risky for Prime Time," available at *http://www.clearwellsystems.com/e-discovery-blog/* (Apr. 20, 2011).

19. Sheila Mackay, "How Dangerous Is Self-Collection in E-Discovery?," available at *http://ediscoverytalk.blogs.xerox.com* (Jul. 30, 2012) (stating that "[w]hile self-collection may seem like a cost-effective way to control costs, the risks of relying on employees to self-collect have been widely covered" and that Judge Scheindlin's July 2012 opinion in the *National Day Laborer Organizing Network* case "comes on the heel of a series of cases demonstrating why companies ought not to rely on their employees to collect data once a legal hold has been implemented").

20. Ralph Losey, "Judge Scheindlin Issues Strong Opinion On Custodian Self-Collection," available at *http://www.law.com* (Jul. 17, 2012).

21. Gonsowski, *supra* n.18; *see also* McKay, *supra* n.16; Shook, *supra*, n.16.

22. *See*, *e.g.*, Nat'l Day Laborer Org. Network v. US Immigration & Customs Enforcement Agency, 2012 U.S. Dist. LEXIS 97863, No. 10-CV-3488 (S.D.N.Y. Jul. 13, 2012); Comm. of the Univ. of Montreal Pension Plan v. Banc of Am. Secs, LLC, 685 F. Supp. 2d 456 (S.D.N.Y. 2010).

23. *See*, *e.g.*, *Green*, 2011 WL 806011at *9; *Nat'l Day Laborer*, 2012 U.S. Dist. at *46, .U.S. Dist. LEXIS 97863, No. 10-CV-3488 at *46 (S.D.N.Y. Jul. 13, 2012).

24. Custodians may be asked to provide input at several stages of the litigation process—from preservation and early case assessment through collection and even review and production. The standards applicable at each stage will vary accordingly. For example, at the preservation stage a custodian may be asked to identify all documents potentially relevant to any aspect of anticipated litigation, whereas at the collection stage the custodian may be asked only to identify documents responsive to specific discovery requests. For simplicity, this article refers to the standard to be applied by custodians in self-collection simply as "potential relevance."

25. *Compare* Qualcomm Inc. v. Broadcom Corp., No. 05cv1958-B, 2008 WL 66932, at *9 (S.D. Cal. Jan. 7, 2008) (describing custodial self-identification of documents, and stating that "attorneys and clients must work together to ensure that both understand how and where electronic documents, records and emails are maintained and to determine how best to locate, review, and produce responsive documents. Attorneys must take responsibility for ensuring that their clients conduct a comprehensive and appropriate document search.") *with Jones*, 2010 WL 2106640 at **3-4 (describing custodians' efforts to "cull out relevant documents" and "print[] out" potentially relevant emails).

26. *See* Fed. R. Civ. P. 26(b)(2)(C) (proportionality), 26(g) (reasonable inquiry); Cache La Poudre Feeds, LLC v. Land O'Lakes, 244 F.R.D. 614, 628 (D. Colo. 2007) (discussing company's obligation to "undertake a reasonable investigation to identify and preserve relevant materials"); Comm. of the Univ. of Montreal Pension Plan v. Banc of Am. Secs, LLC, 685 F. Supp. 2d 456, 461 (S.D.N.Y. 2010) ("In an era where vast amounts of electronic information is available for review, discovery in certain cases has become increasingly complex and expensive. Courts cannot and do not expect that any party can meet a standard of perfection.").

27. The Sedona Principles: Second Edition, Best Practices Recommendations & Principles for Addressing Electronic Document Production (The Sedona Conference Working Group Series, 2007), Principle No. 6; *see also* Kleen Products LLC v. Packaging Corp. of Am., 2012 WL 4498465 at *5,No. 10 C 5711 (N.D. Ill. Sept. 28, 2012) (quoting Sedona Principle No. 6); *Cache La Poudre*, 244 F.R.D. at 628 (same).

28. Suntrust Mortgage, Inc. V. AIG United Guaranty Corp., 2011 U.S. Dist. LEXIS 33118, No. 3:09-cv-529 (E.D.Va. Mar. 29, 2011).

29. *Id*. at **8-9.

30. *Id*. at *16.

31. *Id*.

32. Jones v. Bremen High School District 228, 2010 WL 2106640, *1, *7 (N.D. Ill. May 25, 2010).

33. *Id*. at *1 (emphasis added).

34. *Id*.; *see also* Northington v. H & M Int'l., 2011 WL 662727, No. 08 C 629, at *17 (N.D. Ill. Jan. 12, 2011) (imposing sanction of adverse inference where defendant relied on custodians who allegedly had engaged in misconduct at issue to search their own files for relevant documents).

35. Indeed, most organizations ultimately have no choice but to presume at some level that its employees are honest and forthright because unless they can afford CIA-grade security systems (and as Aldrich Ames, Robert Hanssen, and others have shown, possibly not even then), even a scorched-Earth, image-the-universe approach to document collection likely cannot prevent an employee who is hell-bent on concealing information from permanently deleting files; storing data on an undisclosed device; or even secretly shredding hard copy records or simply hiding a file folder out of sight.

36. Green v. Blitz U.S.A., Inc., No. 2:07-cv-37, 2011 U.S. Dist. LEXIS 20353, *4 (E.D. Tex. Mar. 1, 2011).

37. *Id*. at *4.

38. *Id*. at *6, *9.

39. *Id*. at *6 n. 4.

40. *Id*. at *6.

41. *Id*.; *see also* Comm. of the Univ. of Montreal Pension Plan, 685 F. Supp. 2d at 473 n.68, 496 (imposing sanctions of adverse jury inference and monetary sanctions on plaintiffs after defendants "demonstrated that most plaintiffs conducted discovery in an ignorant and indifferent fashion" and noting that "attorney oversight of the process [of preservation and collection], including the ability to review, sample or spot-check the collection efforts is important."); Phillip M. Adams & Assocs., LLC v. Dell, Inc., 621 F. Supp. 2d 1173, 1194 (D. Utah 2009) (granting in part plaintiff's motion for sanctions against defendant, which failed to produce relevant documents also produced by third

parties, and criticizing defendants' document collection practices, which "place operations-level employees in the position of deciding what information is relevant to the enterprise and its data retention needs."); Pass & Seymour Inc. v. Hubbell Inc., 255 F.R.D. 351 (N.D.N.Y. 2008) (imposing sanctions when outside counsel gave only a "modicum" of guidance regarding client's self-collection); Cache La Poudre Feeds, LLC v. Land O'Lakes, 244 F.R.D. 614 (D. Colo. 2007) (imposing monetary sanctions of $5,000 plus court reporter fees and transcript costs associated with deposition of defendant's general counsel, who testified that, following custodial self collection, he "simply accepted whatever documents or information might be produced by Land O'Lakes employees" but failed to supervise or test these collections); Coleman (Parent) Holdings Inc. v. Morgan Stanley, Inc., 2005 WL 674885 (Fla. Cir. Ct., Mar. 23, 2005) (holding that Morgan Stanley, which had relied on custodial self-collections, had failed to preserve and produce documents and that outside counsel had failed to supervise collections); Wachtel v. Health Net, Inc., 239 F.R.D. 81 (D.N.J. 2006) (imposing sanctions on defendant who relied on business people within the company to collect documents that they determined to be responsive and without supervision of counsel).

42. Nat'l Day Laborer Organizing Network v. US Immigration & Customs Enforcement Agency, No. 10-CV-3488, 2012 U.S. Dist. LEXIS 97863 (S.D.N.Y. Jul. 13, 2012).
43. *Id*. at *3.
44. *Id*. at *4.
45. *Id*. at *7.
46. *Id*. at *46.
47. *Id*. at *46, *citing Comm. of the Univ. of Montreal Pension Plan*, 685 F. Supp. 2d at 473 n.68.
48. *Id*. at *7.
49. *Id*. at *40.
50. *See*, *e.g.*, The Sedona Conference, "The Sedona Conference Cooperation Proclamation," 10 *Sedona Conf. J.* 331 (2009); The Sedona Conference, The Sedona Conference Cooperation Guidance for Litigators & In-House Counsel (2011), available at *https://thesedonaconference.org/publication/The%20Sedona%20Conference%20Cooperation%20Guidance%20for%20Litigators%20%2526%20In-House%20Counsel* ("[D]ialogue and cooperation (rather than unilateral action or intervention) is likely to lead to a collection plan that will find the most important discoverable information, possibly from fewer sources, and likely more quickly and less expensively, and with less uncertainty.").
51. *See*, *e.g.*, *Qualcomm*, 2008 WL 66932 at *9 (attorneys must ensure that custodians understand discovery requests).
52. *See* Peter Kiewit Sons', Inc. v. Wall Street Equity Group, Inc., No. 8:10CV365, 2012 WL 1852048, at *21 (D. Neb. May 18, 2012) (where employee, who stated that she was not the most "computer literate" person employed by defendants, was instructed to create backup of server, court admonished that "she undoubtedly lacks the experience or training to ensure that all files and other information from Server 1 were preserved in the transfer").
53. *See* Achieving Quality, *supra*, at 14-15 ("Quality control processes employed prior to the review of ESI are an essential element to demonstrate the 'reasonableness' of a party's discovery efforts.... Parties using a well-designed discovery methodology should be able to account for all of the electronic information they collect (as well as identify the ESI they did not collect).... Without [effective quality control processes], parties are more vulnerable to potential challenges related to omission of potentially relevant data.").
54. *See* Achieving Quality, *supra* n.14, at 15 ("More advanced technologies have emerged that employ complex algorithms for ESI filtering and organization and, in some cases, may be useful at the collection stage.").
55. *Id*. at 22-23.
56. *See Peter Kiewit Sons'*, 2012 WL 1852049 (following plaintiff's discovery motion, defendants required to engage forensic expert and pay for forensic expert to conduct document searches).
57. *See* Comm. of the Univ. of Montreal Pension Plan, 685 F. Supp.2d at 461.
58. *National Day Laborer*, 2012 U.S. Dist. LEXIS 97863, at *1; *see also* Achieving Quality, *supra* n.14, at 15-16 ("Best practices also call for clear documentation of what was done and not done.").
59. The comments to Sedona Principle No. 6 recognize the importance of documenting and validating collection procedures:

> *All collection processes should be accompanied by documentation and validation appropriate to the needs of the particular case.* Well-documented collection and production procedures enable an organization to respond to challenges—even those made years later—to the collection process, to avoid overlooking electronically stored information that should be collected, and to avoid collecting electronically stored information that is neither relevant nor responsive to the matter at issue. *The documentation of the collection process should describe what is being collected, the procedures employed and steps taken to ensure the integrity of the information collected.*

The Sedona Principles: Second Edition (2007), Cmt. 6.e (emphasis added).