E-Discovery

SHUTTERSTOCK

Big Data And Artificial Intelligence: Implications for E-Discovery's Future

BY MARK S. SIDOTI AND LUIS J. DIAZ

'Big Data"—extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions—is raising new challenges and ethical concerns in litigation. Many corporate legal departments are exploring new technologies, including sophisticated data processing and technology-assisted review tools. As this technology evolves, all signs point to the continued emergence of data evaluation systems based on artificial intelligence (AI) that will have the ability to analyze massive data stores, and even deduce patterns of human behavior, at a fraction of the current cost.

Big Data is increasingly prevalent in the age of the "Internet of Things," with information that is potentially relevant to litigation, and thus targeted in e-discovery, increasing exponentially. Electronically stored information (ESI), which overtook paper discovery years ago, now involves much more than the emails and documents found on enterprise-connected computers. It also derives from social media sites like Twitter, Facebook and LinkedIn, standalone data storage and Internet-connected devices such as smartphones and automobile "black boxes," and new-to-market "smart" household devices like thermostats, security systems and AI-driven home personal assistance devices like Amazon Echo, all of which store data locally and/or on the cloud.

The sheer volume of this data is overwhelming, with worldwide storage of digital information now estimated to be around three zettabytes (one zettabyte is equivalent to 152 million years of high-definition video). With mountains of data accumulating each minute, it is not uncommon to see a complicated litigation matter that involves a terabyte or more of data. In short, Big Data is here to stay, making it difficult, if not impossible, in many cases to conduct cost-effective searches to identify and segregate relevant data for e-discovery purposes without the use of AI-enabled technology. The legal industry has responded by implementing a set of increasingly sophisticated software tools. This trend is expected to continue as software continues to incorporate AI to deal with Big Data in e-discovery.

These **AI-based legal assistants** are expected to make e-discovery **more accessible and affordable** to individuals and mid-sized companies in litigation, **thus enhancing access to justice.**

AI-Enhanced Review

Not so long ago, e-discovery involved primarily the use of keywords that attorneys and consultants would utilize to run Boolean searches against potentially discoverable data sets and newer, "passive" analytic technologies like concept searching and clustering. However, given the exponential increase in data volume, human analysis (linear review) and even so-called "unsupervised machine learning" alone is no longer practical for many lawsuits and investigations. Today, the state-of-the-art in e-discovery is AI-assisted predictive coding. This technology generally uses small "seed sets" of pre-selected relevant documents and the judgment of "subject matter expert" (SME) reviewers to "teach" the AI system to recognize patterns of relevance in the larger document set and rank documents accordingly. The goal of training the system is to ultimately allow the computer—with varying degrees of continuous human input—to accurately predict relevance for the remaining documents in the larger data set. This "active machine learning" process is interactive and iterative, with the most recent studies showing that cycles of continuous reviewer and SME feedback trains the system to more accurately and efficiently select relevant documents.

The research from several studies confirms that predictive coding through a continuous, active machine learning process is indeed faster and less expensive and, most importantly, more accurate than earlier, more labor-intensive methods of identifying relevant evidence in large data sets. See Gordon V. Cormack and Maura R. Grossman, "Evaluation of machine-learning protocols for technology-assisted review in electronic discovery," Proceedings of the 37th International ACM SIGIR Conference on Research & development in information retrieval. ACM, 2014. In an earlier study by RAND Corporation titled "Where the Money Goes: Understanding Litigant Expenditures for Producing Electronic Discovery," the authors concluded that "predictive coding has the potential to lower the cost of unwieldy e-discovery processes by reducing the number of documents requiring human review." That study was based on a review of 57 case studies from eight large corporations and an analysis of the relevant research on electronic discovery review processes, including attendant costs.

The courts have generally welcomed predictive coding. For example, in *Global Aerospace v. Landow Aviation*, No. 61040 (Loudoun County, Va. Cir. Ct. April 23, 2012), the trial judge allowed the use of predictive coding by defendants to review over two million documents over the requesting plaintiff's objections. In an exhaustive memorandum, the defendants successfully argued that predictive coding was "capable of locating upwards of seventy-five percent of the

» Page 10

MARK S. SIDOTI *is chair of the Gibbons e-discovery task force.* LUIS J. DIAZ *is a director in the firm's intellectual property department.*

Strategies for Understanding Your Data Before a Meet-and-Confer

BY GABRIELA BARON

When your clients ask, "What do I need to know before a meet-and-confer?" the answer is simple: Know the data.

Rule 26(f) of the Federal Rules of Civil Procedure (FRCP) states that, in most cases, parties "must confer" to, among other things, "develop a discovery plan." Where required, these conferences provide an opportunity for both sides to craft a thoughtful and proportional approach to discovery. Coming prepared can yield better results in several important areas, from identifying issues with the use of certain types of electronically stored information (ESI) maintained by the parties to deciding limitations around discovery and the optimal document review approaches.

On the surface, being familiar with your client's data seems routine and obvious, but it can be fraught with missteps. A thorough understanding of where the technical minefields are, how they relate to your e-discovery strategy and how to avoid them, enables legal teams to approach meet-and-confers with greater confidence, accelerate the e-discovery process and reduce costs. The meet-and-confer is the parties' best chance to set a reasonable, appropriate technical scope for discovery, which can avoid unnecessary costs for the clients and delay resolution of the dispute.

The three scenarios described below can result from unpreparedness; solutions toward effective conferences are also provided.

(1) You got blindsided at the meet-and-confer because you didn't identify all possible sources of custodial data. As a result, you agreed to an order with aggressive timelines that will allow expansive—and unnecessarily expensive—searches of your client's extensive data repositories, many of which contain duplicate documents.

Solution: Before you develop your e-discovery strategy, understand your client's data landscape (where data is maintained, how easily it is accessed (in terms of both costs and time), and redundancies that may exist among data stores).

Understanding the IT and data landscape, especially around ESI, can help you answer basic technical questions before you inadvertently commit to an onerous discovery process: What is needed to get data from the potential custodians? What is the time period pertinent to the matter? What types of files have been created and where are they stored; mobile devices, back-up tapes, archives and cloud-based data?

This information gathering process can guide legal teams as to a number of other technical items that are important to know before conferring, including:

• Determining whether forensic images of hard drives are necessary and whether any data may be deemed not reasonably accessible under the FRCP.

• Better estimating review times by knowing the volume of email and file types that require special collection, processing and review, such as social media posts, spreadsheets and presentations.

• Determining an optimal list of keywords to propose to the opposing counsel, while allowing collections specialists to advise on the safest ways to preserve metadata and rectify any problems before applying search terms. (The effectiveness of searches varies with different data types and faulty metadata can impede the searchability of documents.)

(2) You brainstormed a series of keywords reflecting the issues in the case, but didn't test them against the document corpus

» Page 11

GABRIELA BARON *is senior vice president at Conduent Legal and Compliance Solutions, where she oversees global business development activities and major client accounts. She may be reached at Gabriela.Baron@conduent.com.*

Benefits and Pitfalls Of Outsourcing Email Services

BY ROBERT A. BANNER AND SEAN SCUDERI

In recent years, companies, large and small, have moved away from maintaining their email servers in-house and have outsourced these services to third-party email service providers, also known as email hosting providers. Email hosting providers offer companies a wide array of appealing services, and they are growing in popularity. They assume management responsibilities for companies' emails, including maintenance, security, data storage, and backup. Email hosting providers offer counsel attempting to obtain documents an additional source of recovery and offer companies seeking to maintain documents a chance to customize retention policies to maximize compliance obligations. The purpose of this article is to highlight these opportunities and potential pitfalls for counsel who may not be aware of them.

Obligation to Produce

With the increased use of email in all areas of business, the volume of electronically stored information (ESI) generated by a company is tremendous. Email hosting providers can help keep a company's ESI organized, backed up, and archived in a secure and accessible manner. However, in the event a company reasonably anticipates litigation, it must remember one important principle: The ESI in the possession of its email hosting provider must be preserved and relevant portions must ultimately be produced.

Although New York courts have not expressly held that companies are obligated to produce those documents in the possession of their email hosting providers, it is a near certainty, based on the existing case law, that courts would compel such production if challenged. It is foolhardy to think that a company could skirt its obligations to preserve ESI by simply outsourcing its email services to an email hosting provider.

Every litigator has received discovery demands requesting documents in his or her client's "possession, custody and control," a phrase that comes directly from CPLR 3120. The Court of Appeals has interpreted the "possession, custody and control" standard contained in the CPLR to mean constructive possession rather than strictly actual possession.[1] This interpretation allows for "discovery from parties that had practical ability to request from, or influence, another party with the desired discovery documents."[2] Based upon this decision, it is clear that any New York judge would hold that a party has an obligation to preserve and produce ESI in the possession of its email hosting provider.

Choosing a Provider

When a company chooses to outsource its emails, it should be sure to choose an email hosting provider that best serves its needs, especially in the area of document retention and destruction. The most important information to understand in this regard is: (1) how often are emails backed up or archived, and (2) how long are backed up and archived emails preserved. Each email hosting provider may have different protocols, and certain email hosting providers allow companies to customize these services. We recommend companies consult with their IT professionals to understand the other practicalities involved, such as whether all ESI is backed up or just select portions and whether certain information is preserved for different periods of time.

Rather than just focus on the ease of services and costs, companies should make sure that satisfactory preservation policies are in place. Some factors that may impact how often ESI should be backed up and how long it should be preserved are industry regulations, the lifespan of business transactions or projects, contractual provisions requiring document retention, or potential threats of litigation. Companies in litigious industries should inquire as to whether the email hosting provider is experienced with handling litigation hold notices and what mechanisms are in place to handle them (along with the associated additional costs). Customization options should continue to increase as the email hosting providers become even more sophisticated.

Understand Policies

It is critically important that all employees understand their employer's internal document retention and destruction policies as well as those of the company's email hosting provider. A clear set of protocols for document retention and destruction will allow employees to understand what types of preservation and deletion activities should be upheld and avoided. Specifically, employees should be able to answer the following questions:

When a client outsources its emails to an email hosting provider, make sure the client understands that it is **obligated to produce all relevant ESI** in the possession of its email hosting provider.

• Are there mechanisms in place to back up ESI? If so, how often are backups performed and how long are the backed up files retained?

• Are there policies in place preventing employees from deleting emails or other ESI in violation of company policies?

• Is ESI routinely deleted after a certain period of time?

• Are their planned programs and resources in place to implement a hold on existing document retention and destruction policies if needed to preserve ESI?

• If an employee inadvertently or impermissibly destroys ESI, what steps should be taken and who should be notified?

Companies should take appropriate measures to ensure their document retention and destruction policies are clearly disseminated to employees. This education is best conveyed through training programs. If the policies are ever changed, a memo explaining the changes in plain English should also be circulated to employees. The employer can also host an informational seminar for his or her employees with the help of its IT professional or even representatives from the email hosting provider. Companies are also encouraged to test out the document retention and destruction policies to have a better understanding of how ESI is preserved and how

» Page 10

ROBERT A. BANNER *is a partner and* SEAN SCUDERI *is an associate at Ingram Yuzek Gainen Carroll & Bertolotti.*

Can **Tests for Spoliation** in NY State, Federal Courts **Be Reconciled?**

BY SAMANTHA V. ETTARI

SHUTTERSTOCK

For many years the sanctions available for the spoliation of electronically stored information (ESI) were largely similar in both the New York federal and state courts. New York state court decisions frequently tracked the federal common law spoliation analysis, most notably set out in the Southern District of New York's *Zubulake v. UBS Warburg* and *Pension Comm. of Univ. of Montreal Pension Plan v. Banc of Am. Secs.* line of cases.[1] This analysis allowed for severe—and sometimes case-terminating—sanctions, such as adverse inference instructions, dismissal of claims or counterclaims, or outright dismissal of actions, for both grossly negligent or intentional spoliation. However, in the past year, with the passage of the December 2015 amendments to the Federal Rules of Civil Procedure (the Rules), the measure for spoliation and accompanying sanctions in New York state and federal courts has diverged.

Under the Federal Rules

Among the December 2015 amendments to the Rules was a significant rewrite of Rule 37(e), which addresses sanctions available for the failure to preserve ESI. Under the revised rule, sanctions are only available if ESI "that should have been preserved in the anticipation or conduct of litigation is lost because a party failed to take reasonable steps to preserve it, and it cannot be restored or replaced through additional discovery[.]" The amendment's most notable departure from certain jurisprudence developed within the Second Circuit concerns remedies available for negligent spoliation, in contrast to the remedies available only for intentional spoliation. Neither set of remedies can be reached now unless the ESI is actually lost—entirely irretrievable from another source or party.

If the preservation obligation and loss requirements are satisfied, negligent spoliation requires a further finding of prejudice. If those criteria are satisfied, the court may deploy a host of sanctions, but none can be case-terminating. The committee notes list the sanctions appropriate for non-intentional spoliation, which include evidence preclusion to offset prejudice; presentation of evidence or argument to the jury regarding the loss of information; and/or jury instructions that would assist in evaluation of that evidence or argument (distinguishable from an adverse inference instruction). Monetary sanctions such as costs and legal fees are also available.

If the preservation obligation and loss requirements are satisfied, negligent spoliation requires a further finding of prejudice. If those criteria are satisfied, **the court may deploy a host of sanctions, but none can be case-terminating.**

If, however, the court finds that a party "acted with the intent to deprive another party of the information's use in the litigation," prejudice is presumed and the court may choose among more severe sanctions, which include a presumption that lost information was unfavorable (in the context of a dispositive motion or bench trial); instruction to the jury that it may or must presume the information was unfavorable; or dismissal of the action or entering of a default judgment. The latter remedy has been utilized to dismiss or bar specific claims and counterclaims, as well. The committee notes emphasize that Rule 37(e) bars these sanctions where the requisite intent is lacking. This is a direct change from pre-amendment common law that allowed for adverse inference sanctions upon finding gross negligence.[2] Moreover, the notes indicate that if the sanctions described in the Rule are too extreme for the situation, lesser sanctions may be awarded, noting that the "remedy should fit the wrong."

In the year following the amendment, the published decisions from the New York federal district courts addressing ESI spoliation under the revised Rule 37(e) were relatively few in number. Some courts continued to look to the common law instead, where the matter or motion was pending prior to the amendment.[3] Others proceeded under the revised rule, even where the litigation pre-dated the amendment—in part because the amendment could be viewed as more lenient toward the spoliator.[4] Those courts that proceeded under the revised Rule have hewn closely to its instructions concerning "loss" and the availability of sanctions for differing levels of culpability, at times even imposing more lenient sanctions in the face of intentional conduct.

The Second Circuit applied amended Rule 37(e) in *Mazzei v. Money Store* and affirmed the trial court's decision not to give an adverse inference instruction where the loss of relevant ESI was not intentional. 656 F. App'x 558 (2d Cir. 2016). And, in *Best Payphones v. City of New York*, the court closely analyzed whether the ESI was "lost," holding that since the movants "did not attempt to retrieve copies of the emails, or the information that was in the emails," from third parties, "which would have cured any violation under Rule 37(e)," there could be no finding of loss and therefore no sanctions, despite negligent failure to preserve ESI. Only reasonable attorney fees and costs were awarded because some responsive documents were later produced as a result of the spoliation motion. 2016 WL 792396, at *5 (S.D.N.Y. Feb. 26, 2016). In *Cat3 v. Black Lineage*, the first decision in the Southern District to interpret and apply Rule 37(e), the court closely applied each of the Rule's requirements and held that relevant ESI, subject to preservation obligations, had been altered intentionally in such a way as to render it "lost," making Rule 37(e)(2) sanctions available. But the court rejected the "drastic sanctions" of dismissal or adverse inference instruction, and instead precluded the spoliator from relying on altered versions of ESI and awarded costs and reasonable attorney fees.[5] Finally, in *Feist v. Paxfire*, the court found that the plaintiff had not intentionally spoliated browser history because she routinely cleaned the hard-drive and the computer had crashed (although the court noted that routine cleaning should have been suspended as part of ongoing preservation obligations); without intent, the court refused to award the extreme sanction of dismissal. The court instead precluded plaintiff from arguing for an award of certain types of damages, for which ESI relevant to the damages analysis had been lost. 2016 WL 4540830 (S.D.N.Y. Aug. 29, 2016).

While the developing case law in the Second Circuit suggests that litigants who negligently or inadvertently spoliate relevant ESI may take comfort that the most severe sanctions will not be available to their adversaries, the same cannot necessarily be said in New York state courts.

Under New York Common Law

Just days after the federal Rule 37(e) amendment took effect, the New York Court of Appeals handed down a seminal decision on spoliation in *Pegasus Aviation v. Varig Logistica S.A.*, 26 N.Y. 3d 543 (2015). The court held that the appellate division had erred in reversing a sanction order imposed by the trial court for the loss of ESI as a result of the failure to implement a litigation hold and multiple computer crashes. The trial court had held that the failure to issue a hold amounted to gross negligence, presumed relevance of the irretrievable ESI, and awarded sanctions in the form of an adverse inference instruction and striking of defendant's answer. In endorsing the trial court's order, the Court of Appeals reiterated the existing New York standard that "adverse inference charges have been found to be appropriate even in situations where the evidence has been found to have been negligently destroyed," implicitly rejecting any incorporation of or movement toward the delineation between sanctions options based on intentional and non-intentional spoliation in the then-two-week-old amendment to Rule 37(e).

The trial and appellate courts in the state accordingly have continued to award more severe sanctions than would be available in federal court in the absence of intentional spoliation. For example, in *Arbor Realty Funding v. Herrick, Feinstein*, the First Department modified the sanction of dismissal, replacing it with an adverse inference instruction and monetary sanctions instead, for the grossly negligent spoliation of ESI as a result of the failure to issue a timely litigation hold, preserve additional relevant custodians' ESI, and suspend routine data destruction, including back-up tape recycling. 140 A.D.3d 607 (1st Dep't 2016). Similarly, in *Cioffi v. S.M. Foods*, the Second Department affirmed the sanction of an adverse inference instruction for the spoliation of ESI even though plaintiffs had not demonstrated the spoliation was "willful rather than merely negligent." 142 A.D. 3d 520, 526 (2d Dep't 2016). In *Ferrara Bros. Bldg. Materials v. FMC Constr.*, the trial court awarded an adverse inference instruction at trial for the loss of ESI as a result of the replacement of certain of defendants' computers during the pendency of litigation, holding that the sanction was "sufficient to strike a balance between the need to ameliorate any prejudice [arising] from the destruction" and "the absence of demonstrable willfulness on the defendants' part." 2016 WL 6583995, at *4 (Sup. Ct. Queens Cty. March 30, 2016). In modifying a trial court decision to provide for an adverse inference instruction in lieu of striking an answer, the Second Department noted in *Peters v. Hernandez* that "striking a pleading is a drastic sanction to impose in the absence of willful or contumacious conduct." 142 A.D. 3d 980, 981 (2d Dep't 2016). These decisions suggest that outright dismissal of a complaint or the striking of an answer for gross negligence alone is not a readily available remedy in New York state court; however, the severe sanction of an adverse inference instruction may be.

Conclusion

Intentional spoliators are likely to face similarly extreme and potentially case-terminating sanctions in both New York state and federal courts. However, litigants should be aware that negligent spoliators are subject to different standards, which may result in less certainty concerning what exposure a non-intentional spoliator may face if ESI is lost.

1. *Zubulake v. UBS Warburg*, 229 F.R.D. 422 (S.D.N.Y. 2004); *Pension Comm. of Univ. of Montreal Pension Plan v. Banc of Am. Secs.*, 685 F. Supp. 2d 456, 469-70, 496-97 (S.D.N.Y. 2010).
2. See *Pension Comm.*, 685 F. Supp. 2d 456; *Residential Funding v. DeGeorge Fin.*, 306 F.3d 99 (2d Cir. 2002).
3. See, e.g., *Stinson v. City of New York*, 2016 WL 54684 (S.D.N.Y. Jan. 5, 2016) (applying common law, grossly negligent spoliation resulted in sanction of permissive adverse inference instruction).
4. See, e.g., *Cat3 v. Black Lineage*, 164 F. Supp. 3d 488 at 496 (S.D.N.Y. 2016) (amended rule "is much more comprehensive" and "in some respects more lenient as to the sanctions that can be imposed for violation of the preservation obligation"); *Best Payphones, v. City of New York*, 2016 WL 792396, at *3 n.2 (E.D.N.Y. Feb. 26, 2016) (since "the application of the new rule does not create issues of feasibility or injustice, the Court will apply the new rule with respect to the electronic evidence at issue here").
5. *Cat3*, 164 F. Supp. at 501-02. Evidence was subsequently submitted to the court demonstrating that there had been no intentional discovery misconduct, and the sanctions motion was withdrawed. See *Cat3 v. Black Lineage*, 2016 WL 1584011 (S.D.N.Y. April 6, 2016).

SAMANTHA V. ETTARI *serves as e-discovery counsel and is a special counsel in the litigation department at Kramer Levin Naftalis & Frankel.*

Outsourcing

« Continued from page 9

documents may be recovered if deleted. The more knowledge and understanding employees have of these procedures, the less likely preservation issues will arise down the road. Finally, the better the procedures in place, if a glitch occurs, a court is more likely to be sympathetic if some documentation cannot be retrieved.

Litigation Hold Notices

Once a company reasonably anticipates litigation, a duty to preserve its documentation is triggered.[3] The Appellate Division, First Department, has explained that the reasonable anticipation of litigation is such time when defendants are on "notice of a credible probability that [they] will become involved in litigation."[4] Once a company reasonably anticipates litigation, it has a duty to suspend its routine document retention and destruction policies and implement a litigation hold to ensure that all relevant documentation is preserved.[5]

If a company outsources its emails, it is not enough for the company to simply suspend its internal document retention and destruction policies. The company should also provide a litigation hold notice to its email hosting provider and request that all document retention and destruction policies in place regarding the company's emails be placed on hold and all ESI be preserved.

Timing is key when it comes to establishing a litigation hold. If a company is late in implementing the litigation hold and potentially relevant documents are destroyed, the company opens itself up to potential spoliation sanctions. A company's use of an email hosting provider may actually help a company avoid such a scenario, depending on the backup and archiving policies in place. There is likely a period of time where deleted emails can be recovered by the email hosting provider if deleted by an employee. We were involved in a case where the defendant failed to implement a litigation hold and continued to delete emails on a daily basis.[6] The defendant sought to recover the deleted emails from its email hosting provider, but it missed its window of recovery.[7] Sanctions were assessed against the defendant.[8]

Conclusion

When a client outsources its emails to an email hosting provider, make sure the client understands that it is obligated to produce all relevant ESI in the possession of its email hosting provider. It is critical that the client understands the obligations it has in anticipation of litigation also encompass its email hosting provider. Ideally, a company will be engaged in an ongoing dialogue with the email hosting provider, have knowledge of its document retention and destruction policies, and have mechanisms in place to provide a litigation hold notice when necessary. Remember, a client can work with its email hosting provider to establish document retention and destruction policies that best suit its needs. Ignoring its obligation to preserve documents in the possession of its email hosting provider can subject the client to a litany of damaging and avoidable sanctions. Working with the email hosting provider should ameliorate many fears about document retention in the age of emails and allow clients to concentrate on their core business concerns.

1. *No. Mariana Islands v. Canadian Imperial Bank of Commerce*, 21 N.Y.3d 55, 62-63 (2013).
2. Id. (citing *Bank of New York v. Meridien BIAO Bank Tanzania Ltd.*, 171 FRD 135, 146 (S.D.N.Y. 1997).
3. See *VOOM HD Holdings v. EchoStar Satellite*, 93 A.D.3d 33, 38 (1st Dep't 2012); *Zubulake v. UBS Warburg*, 220 F.R.D. 212, 217 (S.D.N.Y. 2003).
4. *VOOM*, 93 A.D.3d at 43.
5. *Zubulake*, 220 F.R.D. at 218.
6. *TIAA Global Invs. v. One Astoria Sq.*, 2016 N.Y. Misc. LEXIS 2419 (N.Y. Sup. Ct. Feb. 22, 2016).
7. Id.
8. Id.

Artificial

« Continued from page 9

potentially relevant documents and can be effectively implemented at a fraction of the cost and in a fraction of the time of linear review and keyword searching." Judicial acceptance of predictive coding will likely expand now that the issue of proportionality in discovery and the attendant costs have been brought in to the spotlight by the recent amendments to Federal Rule of Civil Procedure 26.

And already, divergent approaches to the application of AI to e-discovery are emerging. In his series of related blog posts collectively entitled, "Using Hybrid-Multimodal Methods—Predictive Coding 4.0 and Intelligent Spaced Training (IST)," Ralph Losey argues convincingly that the most effective use of predictive coding through AI-based review systems requires constant guidance and ultimate decision making authority by SMEs throughout the process. Losey describes a workflow in which the AI system is continuously trained by the reviewers through a "positive feedback loop" that "continues until the computers predictions are accurate enough to satisfy the proportional needs of the case." A similar but somewhat competing vision is presented by Maura Grossman, whose research with Gordon Cormack is at the forefront of this field. Grossman posits the application of AI based on a "continuous active learning" methodology, which relies more heavily on the computer's initial relevancy ranking determinations.

There are several companies today offering AI-enabled ESI searching capabilities. Products like Kroll Ontrack's EDR software, Catalyst's Insight Predict, BlackStone's Discovery IQ, and Cognitive Analytics NexLP Story Engine incorporate AI into their platforms to not only assist with standard ranking and coding of large document sets, but also develop searches that will identify each category of documents and use AI-driven concept search, enhanced threading, in-depth filtering, visual analytics and sentient analysis. Even so, these systems represent merely the tip of the iceberg in terms or harnessing the potential of AI in the legal field.

The Next Generation

Just as robotic surgical systems have revolutionized the way doctors approach minimally invasive surgery, AI technology is changing the practice of law, even beyond the ESI search and review process. Future systems will continue to leverage increasingly complex predictive coding algorithms for ESI review and beyond. Within the next few years, AI-enabled systems will include natural language interfaces to improve usability. Thus, a lawyer working on a complex breach of contract matter will be able to call his robotic "assistant" via any telephony device and request in plain English that it find all communications and interactions between two key witnesses occurring within a specified date range, prepare a summary report of the substance, and update the same on a weekly basis as additional data is processed. Similarly, systems like ROSS, which is built on IBM's cognitive computer Watson, is designed to read and understand natural language, develop theories of a case when asked questions and conduct legal research. ROSS is on the market and has already been deployed by several major law firms. With appropriate human input, AI-based legal assistants will continuously "learn" from handling various assignments and gain greater efficiency and accuracy in assisting the lawyer.

Other subject-matter applications that perform automated tasks—called "bots" in the AI world—are being developed to assist with specific legal issues. Capable of checking files, linking related data, and producing results on a 24/7 basis and at a fixed cost, bots can be expected to emerge in the near future to aid lawyers with investigation, discovery, complex fact analysis and research. For example, in a pharmaceutical products liability cases, a bot could be used that mines big data in social media on a global basis to find evidence of otherwise obscure adverse events. In a complex breach of contract matter, a bot could assist with locating discovery and conducting legal research regarding performance obligations by the various parties. Eventually, some predict, bots will do much more than locate and rank relevant documents at computer speeds or conduct simple research. They will be able to identify and extract pertinent issues in the case utilizing a natural language interface that will not depend on specific terms, but rather on patterns of behavior, or concepts, buried in terabytes of data.

Conclusion

AI will revolutionize the field of e-discovery, and to some degree the practice of law, in the years to come. The obvious promise of ROSS and other bots developed using similar systems is that they can provide greater efficiency and cost-savings for common legal tasks, including e-discovery. These AI-based legal assistants are also expected to make e-discovery more accessible and affordable to individuals and mid-sized companies in litigation, thus enhancing access to justice. However, unlike medicine, the law will always remain a social science. While AI can assist practitioners in identifying relevant evidence in a case, it cannot do this without constant training and refinement by competent lawyers, and it will never replace the skill and judgment of a lawyer in utilizing that evidence in advocacy.

Obtaining Discovery From EU After GDPR's Passage

BY CHRISTIAN SCHRÖDER, JEFFREY McKENNA AND RENEE PHILLIPS

In April 2016, the European Union (EU) adopted a major overhaul of its data privacy laws to better address new technologies and provide a more coherent approach across different EU Member States. The new law, known as the General Data Protection Regulation (EU) 2016/679 (GDPR) takes effect on May 24, 2018. It will replace the patchwork of national laws created under the old Directive 95/46/EC with a more unified law directly binding each Member State and threatening significant fines amounting to 4 percent of a company's global turnover for non-compliance.

Significantly, the GDPR includes new provisions addressing litigation-related international data transfers. These new provisions create both new perils and opportunities when personal data must be transferred from the EU to the United States for use in discovery.

SHUTTERSTOCK

Article 48: The Good and Bad

When drafting the GDPR, EU legislators recognized the pressure U.S. authorities often place on companies subject to EU data privacy laws and wanted to send a clear signal that companies should resist such pressure and better respect EU privacy restrictions. The result was Article 48, which notes that "any judgment of a court or tribunal and any decision of an administrative authority of a third country requiring [an entity holding EU data] to transfer or disclose personal data may only be recognizable or enforceable … if based on an international agreement, such as a mutual legal assistance treaty (MLAT) … ." The provision presents an obstacle because most Member States do not have MLATs with the U.S., and even those that exist often do not always cover U.S. pretrial discovery. Article 48 further notes that it is without prejudice to other grounds for international transfers set forth in Chapter V of the GDPR.

Grounds for Transfers

Although Chapter V lists several options for legitimizing international data transfers, Articles 46 and 49 will likely provide the most useful mechanisms for transfers to the United States during discovery. This includes:

• Standard Contractual Clauses (Article 46(3)(a)): The EU's Standard Contractual Clauses (SCCs) offer a way to facilitate data transfers on an as needed basis for smaller companies or one-off data transfers. Indeed, because it is possible for multiple entities to enter into SCC agreements, they are particularly well suited for data transfers related to litigation where a variety of U.S.-based entities may need access to the data, such as an e-discovery vendor, a document review provider, contract attorneys, or multiple law firms. The main limitation to using SCCs is that they can only be used if data is being transferred for reasons considered legitimate under the GDPR. Any such use must therefore be assessed within the context of the GDPR's general principles for data transfers and the limitations set forth in Chapter V. For example, a transfer undertaken for national security reasons would not be acceptable unless there were other legal permissions which either standing-alone or in combination with the SCC permitted the transfer.

• Transfer for the Establishment, Exercise or Defense of Legal Claims (Article 49(1)(e)): In U.S. civil proceedings, including pretrial discovery, this provision may in many instances offer the best justification for data transfers. While under 26(1)(d) of the old directive, Member States such as Germany passed national implementing legislation that narrowly limited the 'legal claims' justification to legal matters pending before a court—thereby excluding pre-trial discovery because it was considered to be between the parties and not truly before the court. Since the GDPR does not need to be implemented by separate national legislation, Article 49(1)(e) will apply directly within each Member State and it doesn't contain any such restriction. As a result, German or other European parties to U.S. discovery requests may be able to use this provision.

• Transfer for Important Reasons of Public Interest (Article 49(1)(d)): While this alternative will not apply to data transfers in a civil proceeding, it could apply to U.S. law enforcement requests. However, the scope is not unlimited. According to Article 49(4), the "important reason" cited must be acknowledged by either the EU or the Member States' laws. Examples where data transfers might be in the interest of both U.S. and EU laws would be transfers needed to combat money laundering; for antitrust proceedings; for requests by financial supervisory authorities; or for the purpose of public health.

The transfer of data for national security interests is not likely to be covered because the GDPR doesn't address the use of personal data for national security. In addition, a mere abstract acknowledgment of a public interest is unlikely to be sufficient. The GDPR sets high thresholds for meeting the principle of proportionality.

• Transfer Based on Consent (Article 49(1)(a)): Consent can justify a transfer of personal data for use in discovery. However, attorneys should be aware of the following limitations:

- Consent must be obtained from the data subject, not merely the entity holding the data.

- Article 49(1)(a) requires that consent be explicit, informed and voluntary. Implied consent is not sufficient. Nor are pre-checked boxes or general consents obtained in the abstract.

- Employee consents are often seen as coerced because employees may not feel free to refuse an employer's request. Only where one can document that the transfer cannot harm the employee and where the employee can truly give voluntarily consent without fear of retaliation will employee consent be an option.

- Consent will be hard to obtain where a company does not have direct relationships with customers or third parties whose data will be transferred.

- Consent is not an option where full disclosure regarding the purpose of the transfer can't be given to the data subject, such as in internal investigations.

The GDPR has fundamentally changed the implementation of procedures, opening new possibilities for their use in discovery and wiping the slate clean of localized interpretations of their predecessors under the old directive.

Thus, while consent is viable in certain situations, in practice there will be many scenarios where it can't be used.

• Limited Transfer of Individual Data in Case of Compelling Legitimate Interests of the Data Transferring Party (Article 49(1)(2)): If all other means noted above for legitimizing a transfer fail, this provision can be used if the following criteria are met:

- the transfer is not repetitive and concerns only a limited number of data subjects;

- the transfer is necessary for compelling, legitimate interests of the data transferring entity that are not overridden by the interests or rights and freedoms of the data subject;

- the transferring entity has assessed all the circumstances surrounding the data transfer and has provided suitable safeguards;

- the relevant data protection authority has been informed of the transfer; and

- the data subjects have been informed of the intended data transfer.

The scope of this exception is not entirely clear and its meaning cannot be discerned by interpreting existing statutes or case law. As a result, use of this provision may entail significant risk for the transferring entity.

While many of these methods for legitimizing a data transfer will seem familiar to experienced practitioners, the GDPR has fundamentally changed the implementation of these procedures, opening new possibilities for their use in discovery and wiping the slate clean of localized interpretations of their predecessors under the old directive. Of course, useful guidance issued under the old directive is also no longer binding, so practitioners must be careful not to assume what worked under the directive still works under the GDPR.

In addition, none of these provisions, including Article 49(1)(e), provide carte blanche permission to transfer data irrespective of how broad the request is or what measures are taken to protect the data. Appropriate measures must still be put in place, and the amount of data transferred should be the minimum necessary to achieve the purpose for which the data is being transferred. Among other things, this means that when a transfer is needed for response to a discovery request or subpoena, the scope of the request will almost certainly need to be narrowed. European law does not accept the broad concept of responsiveness used in U.S. discovery. Requests must be tightly focused on only the information and custodians directly relevant and critical to the matter in question.

What Should Companies Do?

In light of the draconian fines possible under the GDPR, companies should make a careful case-by-case assessment of the basis for transferring data discussed above before transferring any data to the United States for use in discovery, law enforcement matters or internal investigations. One size does not fit all. Companies should also follow the measures recommended by the Sedona Conference and the EU Article 29 Working Party. These will include, among other things, using international treaties for justifying transfers (if so available), entering into the SCCs proposed by the EU Commission, minimizing the amount of data actually transferred, redacting or anonymizing personal data wherever feasible, entering into a strong protective order with provisions directly addressing documents subject to EU data privacy laws, processing and hosting the data in the country of origin if possible, or at the very least, filtering the data heavily in the country of origin before transferring it, and meticulously documenting any steps taken to protect the data subjects' privacy.

Finally, companies must be prepared to accept that there is no such thing as a risk-free data transfer from the EU to the United States. While taking appropriate measures can substantially reduce the risk, particularly of a large fine, the GDPR is new and much about the law's implementation remains unclear. Only time will provide more concrete guidance.

CHRISTIAN SCHRÖDER *is a cybersecurity and data privacy partner,* JEFFREY McKENNA *is a senior e-discovery and privacy attorney, and* RENEE PHILLIPS *is an employment law partner at Orrick in the Düsseldorf, San Francisco and New York offices, respectively.*

Strategies

« Continued from page 9

before agreeing to them at the conference. Now, these keywords are returning a host of non-responsive documents that are driving up review costs.

Solution: Invest time up front in an iterative process to refine your search process.

Counsel often wait to refine their search process until it's too late; usually once they've discovered how simultaneously over- and under-inclusive their keywords are. Budget-conscious approaches to e-discovery require early refinement of the search process.

Search consultants can help counsel perform preliminary searches, using key terms and methodologies, and assess the technical results. These experts examine detailed reports describing each term's contributions to the outcome to pinpoint anomalies, such as unexpected terminology or document features—for example, email footers—that could throw off the search results.

During this step, it is important to talk with custodians to understand their lingo. The most effective search terms reflect what custodians see as important in performing their jobs, including abbreviations, acronyms and slang they use when discussing these issues. Linguistic experts can use these discussions to develop a targeted keyword list that yields richer results by accounting for varied communication patterns, synonyms and jargon.

It can also be beneficial for an e-discovery specialist to accompany counsel to the conference. They can offer valuable technical input into negotiations about the search process and provide examples of search results and data visualizations through use of various search methodologies, such as technology-assisted review (TAR). This provides the technical basis to convince opposing counsel of the validity of your proposed e-discovery protocol.

(3) At the meet-and-confer, opposing counsel surprises you by asserting their client's plan to use TAR.

Solution: Have a basic understanding of the technologies that can pinpoint cost savings and efficiencies for your client based on the corpus.

Model Rule 1.1 of the American Bar Association Model Rules, states: "To maintain the requisite knowledge and skill, a lawyer should keep abreast of changes in the law and its practice, including the benefits and risks associated with relevant technology."

In other words, when presented with possible application of TAR, an attorney must be able to grasp the technologies and methods to understand whether to put forth or reject the use of TAR. With the assistance of e-discovery experts, attorneys can gain a technical understanding of the volume and makeup of the file types in the document collection to determine which analytics strategies will be most effective for the case. For example, a case with a high volume of spreadsheets, database reports or non-searchable files may not be ideal for TAR review or other text-based analytics; however, if the majority of the document corpus contains email and word processing, TAR can be a very effective review strategy.

A holistic data analytics strategy gives counsel the ability to expedite meet-and-confer preparedness by ensuring the team has a strong understanding of the document collection from day one.

Emerging Approaches

What is on the horizon for meet-and-confers in 2017? As data volumes continue to grow in complexity, variety and scope, meet-and-confers are becoming more time-intensive and costly for all parties involved. In addition to the tactical, technical steps discussed above, counsel and their clients are going to increasingly adopt "big data" analytics strategies to "know the data" before litigation commences.

Unlike previous-generation analytics that derive insights from documents within a single case (such as TAR, email threading and relationship analysis), big data analytics are capable of consolidating data from litigation, investigation and regulatory compliance matters hosted across a company's internal and third-party e-discovery platforms. They can amass billions of previously reviewed and classified records into a unified repository to identify which documents are relevant for new cases and predict which have the potential of becoming a legal liability.

These insights are derived by data scientists and subject-matter experts applying a variety of customized algorithms and techniques, including machine learning, statistical learning, text analytics, natural language processing, audio analytics and anomaly detection.

A holistic data analytics strategy gives counsel the ability to expedite meet-and-confer preparedness by ensuring the team has a strong understanding of the document collection from day one, along with the following benefits:

• Automatically identifying documents of interest in a matter by repurposing work product, while saving up to 60 percent or potentially millions of dollars per case by eliminating repeat reviews of the same documents.

• Detecting inconsistently classified documents, thus preventing the risk of overlooking relevant documents or exposing private or other sensitive data.

• Helping determine resources, budget and cost of litigation.

Lessons Learned

Counsel who have failed to get the right people on the team at the right time, including search, linguistic, technology, data science and subject matter experts, often miss early opportunities to streamline discovery and conserve time and resources. They have learned that what happens *before* the Rule 26(f) conference—including implementing big data analytics strategies–is as important to a successful outcome as what happens during the conference. The benefits of investing in properly designing a holistic strategy—to "know your data" and avoid going it alone at a meet-and-confer—far outweigh the risks of lost time and money.