Bloomberg BNA

Digital Discovery & e-Evidence®

BEST PRACTICES

The authors describe various strategies for helping a client company identify the information relevant to understanding the case and the information responsive to an opponent's production requests.

Making a Molehill Out of a Mountain: Tips for Handling Terabytes of Data



BY KIM LEFFERT, MICHAEL FRISCH AND PATRICK GARBE

Scenario

A medium-sized company is a defendant in a putative class action lawsuit. Outside counsel negotiated the scope of the plaintiffs' document requests as much as possible, entered into an eDiscovery agreement, and sent an eDiscovery vendor to the client's headquarters to collect data.

The vendor collects nearly a terabyte of data. The defendant's general counsel would like information regarding ways to manage the costs associated with this large amount of data, including considering various data analytics tools in the document review strategy and using skilled, experienced people who understand how to deploy these tools as part of a defensible process.

Goals

Two of the key goals for companies responding to eDiscovery requests are identifying the information that is relevant to understanding the case and identifying the information that is responsive to the opponent's production requests. The following strategies should help the company accomplish these goals while minimizing the cost and burden associated with managing a large volume of data.

Using early case assessment (ECA) tools and review workflow techniques, the case team may be able to prioritize the review and production, reducing the massive volume of data to a more manageable level, and thereby reducing time and cost.

A side benefit of prioritizing the review is that the most relevant documents are typically reviewed early in the process, which allows for ECA and strategy development.

Identifying Relevant Data Sources

The complexity of today's information systems requires organizations that are responding to requests for electronically stored information (ESI) to carefully identify the sources of potentially relevant data that are associated with particular custodians, as well as non-custodial data sources, such as shared drives and structured databases.

The Federal Rules of Civil Procedure require a producing party or its attorney to ultimately ''certif[y] that

 ISSN 1941-3882

to the best of the person's knowledge, information, and belief formed after a reasonable inquiry'' that each request for production of information ''is complete and correct as of the time it is made.''[1] A party cannot meet this standard, which requires a ''reasonable'' effort to locate relevant or potentially relevant evidence,[2] without a good understanding of the sources of data within an organization.

Proactive Measures. It is highly recommended that organizations understand the differences between these data sources and develop policies and procedures that will guide the collection process in advance.

For instance, custodial data collections primarily target e-mail that is located in a person's account, stored locally, on an e-mail server or in an e-mail archive system. It is advisable to consult with the IT department about non-custodial data sources because they are likely to have ''mapped'' the data locations across their systems.

In addition, counsel should consider conducting interviews with IT personnel and potential custodians in order to identify users' data handling habits and the locations of potentially relevant data.

Some of the key topics to cover during a data preservation and collection interview include:

- E-mail systems and storage;
- Computer hardware and asset tracking;
- Policies and procedures for departing employees;
- Records management and document retention policies;
- Bring your own device (BYOD) policy;
- Use of social media and cloud-based storage;
- Collaborative sites, such as SharePoint; and
- Backup systems.

Custodians. Perhaps the most important step of any collection in terms of controlling data volumes is the selection of the custodians from whom e-mails and other records are ultimately processed and reviewed.

Today, depending on a company's retention policy, a single employee's e-mail inbox may contain many thousands of messages.[3]

If counsel can defensibly limit the number of custodians at the collection or data processing stage, this can be a powerful way to reduce the volume of data with which counsel must ultimately grapple.

Identifying Potentially Responsive Data

Once the data has been collected from the various sources mentioned above, it needs to be processed. How this is done requires forethought and planning in order to efficiently manage and control the document review and production phases.

Some case teams may consider removing data that is unlikely to lead to responsive documents. The most common step is to ''DeNIST'' data during the initial processing in order to remove particular file types, primarily program or system files.[4]

Next, a case team may consider targeted searches to identify non-relevant files, often in the form of music, videos and photos.

Similarly, ''junk'' e-mail, such as daily newspaper reports and newsletters, might be culled prior to the application of search terms in order to minimize instances of false positive hits. By excluding these files, the case team might gain greater insight into the data while reducing the volume of data promoted to attorney review.

Applying Filters. Another way to cull irrelevant material is through the use of date restrictions. By applying date filters, which often are agreed upon as part of the meet-and-confer dialogue with the requesting party, the case team can concentrate on a date-restricted set of documents for review and analysis.

The case team might also consider custodian-specific time limitations. For example, if a custodian only worked in the relevant department for two months, there may be no reason to include e-mail from that person's entire tenure at the company. This initial cut can be performed during processing and excluded from the reviewable data.

Search Terms. Once broad cuts are made, the next step is typically to run search terms against the remaining data. Creating a list of search terms is an iterative process that is often developed through discussions with the client and testing the terms against the database.

Search term hit reports may suggest modifications of certain terms in order to identify relevant documents in addition to minimizing the amount of false hits.

Counsel should consider working closely with eDiscovery experts to create a search term list, as some courts have required expert testimony about the efficacy of search methodology utilized in disputes about the sufficiency of searches.[5]

Consider the Use of Data Analytics Tools

New technologies can make the review process more efficient and can get attorneys' eyes on the key docu-

[1] Fed. R. Civ. Proc. 26(g)(1).

[2] *Zubulake v. UBS Warburg LLC*, 220 F.R.D. 212, 217 (S.D.N.Y. 2003).

[3] According to the Radicati Group, a firm specializing in quantitative and qualitative research on e-mail and information archiving, the majority of the world's e-mail traffic comes from business users, which account for over 108.7 billion e-mails sent and received per day in 2014. Business users sent and received on average 121 e-mails a day in 2014. This number is expected to grow to 140 e-mails a day by 2018. Sara Radicati, PhD, *Email Statistics Report 2014-2018* (The Radicati Group, 2014).

[4] ''NIST'' is an acronym for the National Institute of Standards and Technology, a government agency that maintains a master list of known, traceable computer applications. To ''DeNIST'' means using that master list to identify computer files known to be unimportant system files and remove them from a document collection. *See* http://www.nextpoint.com/blog/denist-ediscovery/ (last visited March 15, 2015).

[5] *See, e.g., Victor Stanley, Inc. v. Creative Pipe, Inc.*, 250 F.R.D. 251, 262 (D. Md. 2008) (holding that defendants failed to demonstrate that the keyword search they performed on text-searchable ESI was reasonable, where defendants identified neither the keywords selected nor the qualifications of the persons who selected them to design a proper search); *Equity Analytics, LLC v. Lundin*, 248 F.R.D. 331, 332-33 (D.D.C. 2008) (requiring the party challenging the search terms to present expert testimony to explain the deficiencies).

  DDEE ISSN 1941-3882

ments faster. Attorneys usually refer to these new tools collectively as ''technology assisted review'' or TAR.

Some TAR tools reduce costs simply by making document reviews more efficient and faster.

For instance, ''concept clustering'' uses software to group e-mails about certain themes. This way, a single reviewer becomes an ''expert'' on an issue or discussion, and is able to quickly recognize documents that are important, or new.

E-mail threading can reduce review volume by showing reviewers only the most ''complete'' e-mail in a long chain, and automatically code its subsidiary parts so they do not need to be individually reviewed. This technology can help avoid the frustration that any young attorney can attest to of seeing the same e-mail chain appear again and again in a review set.

This technology can save large amounts of time, especially in cases where attorneys choose not to ''deduplicate'' e-mails collected across custodians.

The case team might also consider the use of predictive coding during discovery or trial preparation.

Wen contemplating the use of advanced analytics for filtering large data volumes, hiring an appropriate eDiscovery vendor is an important first step.

Predictive coding uses complex algorithms to make relevancy determinations on the entire universe of documents within a database. The computer's decisions are based on a subset of documents that is first reviewed and coded by an attorney or group of attorneys most familiar with the substantive issues in a case. This subset of documents is referred to as a ''seed set.''[6]

Based on the decisions attorneys make when reviewing the ''seed set,'' the computer system learns what is most likely to be relevant or responsive.

Initially, there was some reluctance within the legal community to embrace predictive coding due in part to the ''black box'' nature of the technology and lack of judicial endorsement. In recent years, the use of these technologies has increased and some courts have accepted its use in responding to discovery requests.[7]

Although TAR tools were initially developed and marketed as a means for reducing the first-level attorney review costs, the focus today is trending toward using these tools to improve the evaluation of both documents produced and those received in production.

In addition, data analytics tools like predictive coding can be considered for prioritizing the review workflow; streamlining the second-level review, which is typically performed by outside counsel; and quality-checking the review in order to prepare the documents for production.

Predictive coding and other TAR tools can also save time, and potentially provide better results, during the preparation of witness files for depositions and trial.

Choose Your E-Discovery Partner Wisely

Counsel is advised to select an eDiscovery partner prior to entering the meet-and-confer dialogue. The primary reason for making an early selection is to ensure that your eDiscovery vendor can deliver the services in line with the provisions in the ESI agreement.

For instance, when contemplating the use of advanced analytics for filtering large data volumes, hiring an appropriate eDiscovery vendor is an important first step.

Counsel should keep in mind that not all predictive coding or TAR tools are the same. Counsel is well advised to educate themselves on the technology behind the data processing and their review protocol.

Further, counsel should consider working with a qualified eDiscovery vendor that:

- Can perform forensically sound data collections, process data using defensible workflows and prepare supporting documentation;
- Can make available ECA tools for filtering, searching and developing review strategies;
- Will host the results in a review application that facilitates further analysis; and
- Has the experience and the resources to support the case team and meet discovery deadlines.

In cases with a large volume of data, recurring ''hosting charges'' can become a real burden, especially during a long-running case.

Vendors typically charge a per-gigabyte fee for hosting data. The use of various ECA tools can result in additional hosting charges.

Counsel may want to explore negotiating alternative fee arrangements for processing and/or hosting at the outset.

Document, Document, Document

Whether your opponent's consent is obtained or not, and whatever choices are made for data review, it is important to carefully document them.

Incorporating advance analytics into the collection, review and production phases is relatively new and is still in the process of being fully understood by the legal community. As a result, a degree of skepticism can exist when they are raised during the meet-and-confer dialogue.

Thus, the case team is encouraged to work closely with their eDiscovery provider to create supporting documentation that describes the process. This documentation can be used to replicate the process in future

[6] *See Da Silva Moore v. Publicis Groupe SA*, 287 F.R.D. 182, 183-184 (S.D.N.Y. Feb. 24, 2012).

[7] *Rio Tinto PLC v. Vale S.A.*, 2015 BL 54331 (S.D.N.Y. 2015) (approving parties' use of predictive coding in discovery, and collecting cases approving of the use of the technology); *see also Da Silva Moore*, 287 F.R.D. 182 at 183; *Global Aerospace Inc. v. Landow Aviation, L.P.*, No. CL 61040 (Va. Cir. Ct. April. 23, 2012) (permitting defendant's use of predictive coding ''for the purposes of the processing and production of electronically stored information''); *In re Biomet M2a Magnum Hip Implant Prods. Liab. Litig.*, No. 3:12-MD-2391 (N.D. Ind. April 18, 2013) (finding that Biomet's use of predictive coding satisfied the party's discovery obligations under the Federal Rules of Civil Procedure).

litigation and to explain and defend the process in the event of a challenge.

If Possible, Cooperate

Finally, while it may not always be possible to do so, obtaining ''buy in'' from opposing counsel before key decisions are implemented is a good practice.

Once opposing counsel has blessed a decision or procedure, she will be hard-pressed to argue for increasing the scope of a collection or document review after it has been completed.

For this reason, ''fronting'' major decisions with opposing counsel may ultimately be a cost-saving strategy, as it often prevents an opponent from second-guessing your work and arguing that something must be re-done.

Commentators have noted that while eDiscovery issues continue to be fertile grounds for disputes, judges are doing everything they can to change the ''culture of discovery from adversarial conduct to [one of] cooperation.''[8]

While the current Federal Rules do not specifically mandate a ''duty to cooperate'' (although the proposed amendments to the Federal Rules are moving toward such a duty)[9], all judges expect (and most local rules require) parties to negotiate in good faith about ESI and other discovery issues before approaching the court for help.[10]

Federal Rule of Civil Procedure 1 does call for the ''just, speedy, and inexpensive'' determination of every action—a goal that cooperation will undoubtedly support. Cooperation on these thorny issues will surely reduce costs or at least streamline an unavoidably costly process, and most importantly will keep your judge happy.[11]

For inquiries related to this article, please contact Michael Frisch at mfrisch@mayerbrown.com, Patrick Garbe at pgarbe@mayerbrown.com or Kim Leffert at kleffert@mayerbrown.com.

Learn more about Mayer Brown's Electronic Discovery & Information Governance practice by visiting www.mayerbrown.com or contacting Eric Evans at eevans@mayerbrown, Ethan Hastert at ehastert@mayerbrown.com, Michael Lackey at mlackey@mayerbrown.com or Edmund Sautter at esautter@mayerbrown.com.

[8] *See* Thomas Y. Allman, *Local Rules, Standing Orders, and Model Protocols: Where the Rubber Meets the (E-Discovery) Road*, 19 RICH. J. L. & TECH 8 (2013) at 29, available at http://jolt.richmond.edu/v19i3/article8.pdf (*citing* The Sedona Conference®, *The Sedona Conference Cooperation Proclamation*, 10 SEDONA CONF. J. 331, 331 (2009)).

[9] Rule 1 now provides that the civil rules ''should be construed and administered to secure the just, speedy, and inexpensive determination of every action and proceeding.''

The proposed amendment would provide that the rules ''be construed, administered, *and employed by the court and the parties* to secure the just, speedy, and inexpensive determination of every action and proceeding.''

The purpose of this change was to ''make clear that parties as well as courts have a responsibility to achieve the just, speedy, and inexpensive resolution of every action.'' *See Proposed Amendments to the Federal Rules of Civil Procedure*, Appendix B-13, Committee of Rules on Practice and Procedure of the Judicial Conference of the United States (September 2014), available at http://www.uscourts.gov/uscourts/RulesAndPolicies/rules/Reports/ST09-2014-add.pdf (last visited March 15, 2015).

[10] *See* Allman, *supra* fn. 8 at 27-28. For example, the local rules for the Northern District of Illinois, like many other district's rules, require a party to certify that it has complied with a stringent ''meet and confer'' requirement before filing any discovery-related motion. N.D. Ill. Local Rule 37.2.

[11] *Id.* (noting that ''the judicial enthusiasm for cooperation is widespread and growing'').

Kim Leffert is a counsel in Mayer Brown's Chicago office and a member of the Litigation and Dispute Resolution Practice. One of the founding members of the Firm's Electronic Discovery and Information Governance practice, a significant and growing part of Kim's work involves electronic discovery issues.

Michael Frisch is a senior associate in Mayer Brown's Chicago office and a member of the Litigation & Dispute Resolution practice. Mike's practice focuses on conducting high-profile internal investigations, white collar criminal defense and compliance and complex commercial litigation.

Patrick Garbe is an Electronic Discovery Specialist in the New York office of Mayer Brown LLP's Electronic Discovery Services group and a member of the firm's Electronic Discovery and Information Governance practice. Patrick is a former paralegal with 30 years of experience in discovery services, legal management and paralegal support.

  DDEE ISSN 1941-3882