

# Constructing a Successful E-Discovery Strategy: Foundational Principles and Building Blocks

*Eric A. O. Ruzicka and Kate Johnson\**

## I. INTRODUCTION

Just as a construction project is built with bricks and mortar, a construction litigation strategy is built with documents. And, while the records tell the story, efficiently finding the right documents poses a significant challenge. Construction projects generate an astonishing number of materials. Drawings and specifications, schedules, formal correspondence, meeting minutes, timesheets, photographs, inspection records, project logs and diaries, and e-mails all tell a piece of the story of how a project was built, where it went wrong, and who was responsible.

The proliferation of laptops, smartphones, and tablets on every construction site has facilitated the creation of an even larger number of potentially relevant, potentially case-critical materials, as the generation of electronically stored information (“ESI”) has moved out of the home office and site trailers, and into the hands and pockets of every person on site. As the volume of ESI expands, not surprisingly, the time and expense associated with finding the critical records that support (and weaken) your claims and defenses—and, more generally, complying with discovery obligations by collecting, reviewing, and producing requested documents—grows exponentially.

As the volume of records soars, the legal industry has developed several techniques and technologies to manage discovery: search terms help find relevant items in large document populations; technology-assisted review (“TAR”) and analytic techniques help prioritize materials for review, identify the relevant ones, and exclude the irrelevant. Construction documents, however, frequently evade these methods, which largely rely on searchable written content not available in many typical construction records: photographs, drawings, and schedules frequently do not have sufficient written content to be located or analyzed using these technologies and techniques. Handwritten documents, such as project logs and diaries, generally do not yield machine-

---

\*Eric Ruzicka and Kate Johnson are partners in the Construction and Design practice group at Dorsey & Whitney, LLP in Minneapolis, MN.

readable content, and optical character recognition (“OCR”) programs currently may not provide sufficiently-accurate results to successfully identify relevant documents using search terms or TAR.

To successfully prosecute or defend a construction case, both construction litigators and their clients must understand the current discovery landscape, and the benefits (and drawbacks) of electronic discovery technologies; they provide opportunities to control costs, successfully develop claims and defenses, and gain a competitive advantage. Understanding these issues *before* litigation begins, moreover, allows industry participants to implement litigation-readiness policies and procedures that will help ensure that information that is needed to prove or defend claims is retained, maintained in a logical and accessible fashion, and can be efficiently and effectively collected, should litigation arise.

We first will discuss the current electronic discovery landscape, including the increasing focus on mobile device data and the renewed emphasis on proportionality in discovery. Next, we will explore the benefits and limitations of developing document review and analysis technologies. Finally, we will propose some litigation-readiness, best practices intended to facilitate effective and efficient discovery.

## II. SURVEYING THE CURRENT DISCOVERY LANDSCAPE

In a hallmark of the American legal system, litigants generally are entitled to broad discovery of “any nonprivileged matter that is relevant to any party’s claim or defense . . .”<sup>1</sup> While they understandably want to find the smoking gun in their opponents’ documents, this broad mandate long has steered a collision course with the rapidly expanding universe of electronic documents and data. The tension between this broad discovery mandate and the vast quantity of potentially-discoverable information has resulted in massive discovery expenditures by litigants, while simultaneously spurring dramatic innovation with the electronic discovery industry and an increased focus on proportionality in discovery.

As computers became common in the workplace, electronic data quickly began to dominate all other record forms: by 1999, 93% of information was generated digitally.<sup>2</sup> The amount of data generated globally each year is astonishing: in 2016, 16.1 zettabytes

---

<sup>1</sup>Fed. R. Civ. P. 26(b).

<sup>2</sup>In re Bristol-Myers Squibb Securities Litigation, 205 F.R.D. 437, 440 n.2, 51 Fed. R. Serv. 3d 1212 (D.N.J. 2002).

(or 16.1 billion terabytes) of data.<sup>3</sup> Although document sizes can vary significantly, one recent analysis of records collected for litigation found that a terabyte of data contains between 4,000,000 and 5,000,000 documents.<sup>4</sup> And, the amount of data generated each year is only increasing: by 2025, the International Data Corporation projects that it will be 163 zettabytes.<sup>5</sup>

The continually-growing amount of electronic data presents several challenges for companies, not least of which is the cost to them when litigation arises. A 2015 survey of corporate counsel found that 36% of respondents spent more than \$1 million in the prior year on collection, processing and reviewing electronic data, with 15% spending more than \$10 million.<sup>6</sup>

Two recent issues illustrate the tension between the competing concerns of fulsome discovery and exorbitant discovery costs, and the attempt of the legal system to find some balance. First, as mobile devices have become ubiquitous in both personal and professional environments, the courts have resoundingly required litigants to preserve, collect, and produce relevant data found on them, which can present both financial and logistical challenges. And second, faced with significant concerns regarding the cost of extensive discovery, and the time required to complete it, courts have required that discovery “be proportional to the needs of the case.”<sup>7</sup>

There is often no escaping extensive—and expensive—discovery in construction disputes. However, understanding the legal landscape—including both discovery obligations and opportunities to control the scope of discovery—allows industry participants, and their lawyers, to proactively and strategically manage the process and minimize both costs and risks.

### ***A. The Smoking Gun is in Your Hand: Discoverability of Mobile Devices***

Perhaps no source of electronic data has proven more troublesome in recent years than mobile devices. Mobile device data

---

<sup>3</sup>D. Reinsel, J. Gantz, & J. Rydning, *Data Age 2025: The Evolution of Data to Life-Critical* (April 2017), available at <http://www.seagate.com/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf>.

<sup>4</sup>J. Tredennick, *How Many Documents in a Gigabyte? Revisiting an E-Discovery Mystery* (Aug. 20, 2015), available at <https://catalystsecure.com/blog/2015/08/how-many-documents-in-a-gigabyte-revisiting-an-e-discovery-mystery/>.

<sup>5</sup>Reinsel, Gantz & Rydning.

<sup>6</sup>KPMG, *Managing Electronic Data for Litigation and Regulatory Readiness* (2016), available at <https://assets.kpmg.com/content/dam/kpmg/pdf/2016/02/litigation-survey-2016.pdf>.

<sup>7</sup>Fed. R. Civ. P. 26(b).

(including data such as email, notes, photographs, and text messages) undeniably is discoverable,<sup>8</sup> and litigants increasingly request such data, with the backing of the courts. But, preserving and collecting mobile device data can present significant challenges and costs: devices are lost or updated; data may be stored in the cloud; and employees may use their personal mobile device (or devices) for work. Forensic imaging of a single device can take several hours, and each one can cost several hundred dollars, or more, to image. Development of new devices and technologies can outpace the development of tools to collect data from those devices. But, these devices may hold the critical information that tells your story: the photographs that prove the now-concealed challenges encountered in the field; the text messages that corroborate a key witness's testimony; or the voicemails that provide contemporaneous evidence of an opposing party's admission. Conversely, the consequences of failing to properly preserve, collect, and produce discoverable mobile device data requested by an opposing party can be severe.

Courts have issued spoliation sanctions against parties that failed to preserve and produce text messages—even in the absence of evidence that the lost text messages contained relevant evidence, and where there was no evidence that the failure to preserve them was anything other than negligent.<sup>9</sup> In one such case, the adverse party was permitted to introduce evidence of the failure to preserve the messages and “argue for whatever inference they hope the jury will draw.”<sup>10</sup> The court's ruling was clear: “[a] commercial party represented by experienced and

---

<sup>8</sup>For instance, Federal Rule of Civil Procedure 34(a) permits discovery of “any designated documents or electronically stored information—including writings, drawings, graphs, charts, photographs, sound recordings, images, and other data or data compilations—stored in any medium from which information can be obtained either directly or, if necessary, after translation by the responding party into a reasonably usable form.”

<sup>9</sup>The sanction imposed likely would have been different following the 2015 revisions to the Federal Rules of Civil Procedure. Newly revised Rule 37(e) provides that an adverse inference instruction is appropriate “only upon a finding that the party acted with the intent to deprive a party of the information's use in the litigation.” *See, e.g., Nuvasive, Inc. v. Madsen Medical, Inc.*, 2016 WL 305096 (S.D. Cal. 2016) (vacating an adverse inference instruction based on a party's failure to preserve text messages on a motion for reconsideration based on the revised Rule 37(e), though allowing each side to present evidence regarding the loss of electronically stored information and specifying an instruction that the jury may consider such evidence, along with all other evidence, in making its decision).

<sup>10</sup>*Christou v. Beatport, LLC*, 2013-1 Trade Cas. (CCH) ¶ 78230, 2013 WL 248058, \*14 (D. Colo. 2013).

highly sophisticated counsel cannot disregard the duty to preserve potentially relevant documents . . .”<sup>11</sup>

Nor is it enough to preserve only the substantive content contained on mobile devices. In another recent case, the plaintiff relied on certain photographs taken with her cell phone to prove her claim, and produced printed copies of them in discovery. However, the plaintiff did not maintain digital versions of the photographs, meaning that relevant metadata, including the date and time they were taken, was lost. The court ultimately determined that a permissive adverse-inference instruction, informing the jury that it may, but was not required, to make an adverse inference about the authenticity of the photographs based on a failure to preserve the digital versions, was appropriate.<sup>12</sup>

So, too, have litigants faced sanctions for failure to produce data stored in the cloud<sup>13</sup> and failure to preserve data from personal mobile devices used for work purposes.<sup>14</sup> Notably, in the latter case, the relevant custodians denied using their personal cell phones for work during custodial interviews; however, an analysis of the custodians’ email revealed that work-related emails were sent from personal phones, and certain custodians subsequently admitted in testimony that they used their own cell phones for work-related purposes.<sup>15</sup> By the time this information was uncovered, however, there was no responsive ESI stored on the devices, leading the court to conclude that the failure to identify, preserve, collect or search the devices likely resulted in the destruction of responsive ESI.<sup>16</sup>

With a smartphone in every pocket on a construction site, companies in the construction industry, and the attorneys who represent them, must be prepared for the inevitable need to collect data from mobile devices, and must do so before that data disappears. This preparation must start at the beginning of a project—or before—with document retention policies that address mobile device data. Companies must know if and how their em-

---

<sup>11</sup>Christou v. Beatport, LLC, 2013-1 Trade Cas. (CCH) ¶ 78230, 2013 WL 248058, \*14 (D. Colo. 2013).

<sup>12</sup>Gilley v. Eli Lilly and Co., 2013 WL 1701066 (E.D. Tenn. 2013), report and recommendation adopted, 2013 WL 1694436 (E.D. Tenn. 2013).

<sup>13</sup>Quantlab Technologies Ltd. (BGI) v. Godlevsky, 2014 WL 651944 (S.D. Tex. 2014).

<sup>14</sup>Small v. University Medical Center of Southern Nevada, 2014 Wage & Hour Cas. 2d (BNA) 166346, 2014 WL 4079507 (D. Nev. 2014).

<sup>15</sup>Small v. University Medical Center of Southern Nevada, 2014 Wage & Hour Cas. 2d (BNA) 166346, 2014 WL 4079507 (D. Nev. 2014).

<sup>16</sup>Small v. University Medical Center of Southern Nevada, 2014 Wage & Hour Cas. 2d (BNA) 166346, 2014 WL 4079507 (D. Nev. 2014).

ployees are using their own mobile devices for work, and institute and enforce appropriate Bring Your Own Device (“BYOD”) policies. And, companies must be mindful of whether litigation is reasonably anticipated on a project and, if it is, timely institute a litigation hold that includes mobile devices.

Instructing employees to preserve data on their mobile devices, however, may not be sufficient, especially where employees use their own devices. Individuals are unlikely to know *how* to protect data on their devices from deletion, or protect against lost or broken devices. A litigation hold may be the last thing on an employee’s mind when replacing a device, potentially leaving data behind. And, as projects approach an end—especially troubled projects headed towards litigation—and employees begin to roll off the project, grabbing their cell phones and imaging them may be low on the list of priorities. All of these issues, however, must be proactively addressed to avoid losing data and exposing a company to negative consequences in litigation.

### ***B. Applying the Proportionality Standard***

With this ever-expanding universe of information, is there any hope for the construction industry? The rules governing discovery generally have permitted litigants to “obtain discovery regarding any nonprivileged matter that is relevant to any party’s claim or defense . . .”<sup>17</sup> The 2015 Amendments of the Federal Rules of Civil Procedure took a step towards addressing the continually expanding burden of document discovery by explicitly addressing the requirement that discovery be “proportional to the needs of the case . . .”<sup>18</sup> A proportionality standard also has been adopted by a majority of states.<sup>19</sup> Alternative dispute resolution procedures provide similar guidance; for instance, the AAA Construction

<sup>17</sup>Fed. R. Civ. P. 26(b).

<sup>18</sup>Fed. R. Civ. P. 26(b). The concept of proportionality, of course, is not new, but litigants and the courts have struggled to put proportionality into practice. *See, e.g.,* Laporte and Redgrave, A Practical Guide to Achieving Proportionality Under New Federal Rule of Civil Procedure 26, 9 Fed. Courts L. Rev. 19–39 (2015). Rather than proposing something new, the amended Rule 26 “crystalizes the concept of reasonable limits on discovery through increased reliance on the common-sense concept of proportionality . . . The amended rule states, as a fundamental concept, that lawyers must size and shape their discovery requests to the requisites of a case.” John Roberts, *2015 Year-End Report on the Federal Judiciary* (Dec. 31, 2015), available at <http://www.supremecourt.gov/publicinfo/year-end/2015year-endreport.pdf>.

<sup>19</sup>As of the date of this article, 39 states have adopted a proportionality standard similar or identical to the federal rule: Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Maine, Maryland, Minnesota, Montana, Nevada, New Hampshire, New Jersey, New Mexico, North Carolina, North Dakota, Ohio, Oklahoma, Pennsylvania, Rhode Island, South Carolina, South Dakota,

Industry Procedures for Large, Complex Construction Disputes allow that parties “may conduct such discovery as may be agreed to by all parties,” while empowering arbitrators to establish the extent of discovery “consistent with the goal of achieving a just, efficient and cost-effective resolution . . .”<sup>20</sup>

Judges and arbitrators are empowered—and encouraged, if not expected—to help control the scope of discovery.<sup>21</sup> Yet, determining what constitutes proportional discovery remains a significant challenge. Federal Rule of Civil Procedure 26(b) offers six factors to consider: (1) the importance of the issues at stake in the action; (2) the amount in controversy; (3) the parties’ relative access to relevant information; (4) the parties’ resources; (5) the importance of the discovery in resolving the issues; and (6) whether the burden or expense of the proposed discovery outweighs its likely benefit.

Even with everyone’s best intentions, assessing these factors may be difficult:

[t]he parties may begin discovery without a full appreciation of the factors that bear on proportionality. A party requesting discovery, for example, may have little information about the burden or expense of responding. A party requested to provide discovery may have little information about the importance of the discovery in resolving the issues as understood by the requesting party.<sup>22</sup>

Although the relevant information may be difficult to obtain, parties must be prepared to provide all of the information possible when advocating for their positions on matters relating to proportionality and the scope of discovery:

[a] party claiming undue burden or expense ordinarily has far better information—perhaps the only information—with respect to that part of the determination. A party claiming that a request is important to resolve the issue should be able to explain the ways in which the underlying information bears on the issues as the party understands them. The court’s responsibility, using all the informa-

---

Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, and Wyoming.

<sup>20</sup>American Arbitration Association Construction Industry Arbitration Rules, Procedures for Large, Complex Construction Disputes Rules 4(d) and (e).

<sup>21</sup>Chief Justice Roberts observed, “The 2015 civil rules amendments are a major stride toward a better federal court system. But they will achieve the goal of Rule 1—‘the just, speedy, and inexpensive determination of every action and proceeding’—only if the entire legal community . . . step up to the challenge of making real change . . . Judges must be willing to take on a stewardship role, managing their cases from the outset rather than allowing parties alone to dictate the scope of discovery . . .” Roberts.

<sup>22</sup>*Carr v. State Farm Mutual Automobile Insurance Company*, 312 F.R.D. 459, 467, 93 Fed. R. Serv. 3d 483 (N.D. Tex. 2015), subsequent determination, 2016 WL 269619 (N.D. Tex. 2016).

tion provided by the parties, is to consider these and all the other factors in reaching a case-specific determination of the appropriate scope of discovery.<sup>23</sup>

Whether the renewed focus on proportionality will appreciably impact the cost and burden of discovery remains an open question. However, two things are undeniably true for construction litigants. First, we must be prepared to quickly and effectively gather available information regarding our clients' documents so that we can effectively advocate for an appropriate scope of discovery.

Second, the volumes of data generated during a large construction project, and the types of claims asserted in construction litigation, virtually guarantee that even discovery that is proportional, within the meaning of Fed. R. Civ. P. 26(b), will be extensive. Thus, litigants are well-served to embrace technologies designed to more efficiently and effectively review and analyze the relevant data. As the Committee Notes regarding the 2015 amendments to Rule 26(b)(1) explain,

The burden or expense of proposed discovery should be determined in a realistic way. This includes the burden or expense of producing electronically stored information. Computer-based methods of searching such information continue to develop, particularly for cases involving large volumes of electronically stored information. Courts and parties should be willing to consider the opportunities for reducing the burden or expense of discovery as reliable means of searching electronically stored information become available.<sup>24</sup>

### III. USE OF TECHNOLOGY TO EASE THE DISCOVERY BURDEN

Technology-Assisted Review (“TAR”),<sup>25</sup> including predictive coding,<sup>26</sup> have gained increasing acceptance in recent years. These methods largely make use of a document's content, based on

<sup>23</sup>*Carr*, 312 F.R.D. at 467.

<sup>24</sup>Fed R. Civ. P. 26 advisory comm. nn.

<sup>25</sup>Technology-Assisted Review has been defined as:

“[a] process for Prioritizing or Coding a Collection of Documents using a computerized system that harnesses human judgments of one or more Subject Matter Expert(s) on a smaller set of Documents and then extrapolates those judgments to the remaining Document Collection. Some TAR methods use Machine Learning Algorithms to distinguish Relevant from Non-Relevant Documents, based on Training Examples Coded as Relevant or Non-Relevant by the Subject Matter Experts(s), while other TAR methods derive systematic Rules that emulate the expert(s)' decision-making process. TAR processes generally incorporate Statistical Models and/or Sampling techniques to guide the process and to measure overall system effectiveness.”

Grossman and Cormack, *The Grossman-Cormack Glossary of Technology-Assisted Review*, 7 Fed. Courts L. Rev. 1, 32 (2013).

<sup>26</sup>Predictive coding has been defined as “[a]n industry-specific term generally used to describe a Technology-Assisted Review process involving the use of



extracted text or optical character recognition (“OCR”), and machine learning algorithms to analyze, classify, and/or prioritize documents for review. TAR techniques bypass many human assumptions and allow the documents to “speak for themselves,” based on their content.

Since 2012, when predictive coding first was approved by a court in *Da Silva Moore*,<sup>27</sup> “the case law has developed to the point that it is now black letter law that where the producing party wants to utilize TAR for document review, courts will permit it,”<sup>28</sup> and may even encourage<sup>29</sup> or order it when resolving disputes.<sup>30</sup> TAR is also gaining traction in jurisdictions overseas.<sup>31</sup> But, “computer-assisted review is not a magic, Staples-Easy-Button, solution appropriate for all cases. The technology exists and should be used where appropriate, but it is not a case of machine replacing humans . . .”<sup>32</sup>

While no single button can resolve all discovery-related issues, various technologies can assist attorneys both in complying with discovery obligations and in identifying the documents necessary to support their claims and/or defenses. These methodologies range from simple tools that identify the volume of data in a collection to predictive coding algorithms that can code millions of documents, based on an initial human review of a small sampling of those records. By engaging with these technologies at all stages of a case, attorneys can develop an early understanding of a case, craft a discovery strategy, and find the key documents necessary to building a litigation strategy.

#### **A. Early Case Assessment and Discovery Negotiations**

There may be little time in the early phases of litigation to collect and process documents and begin assessing them before negotiations over the scope of discovery begin and discovery requests come due. To the extent that there is time, however,

---

a Machine Learning Algorithm to distinguish Relevant from Non-Relevant Documents, based on Subject Matter Expert(s)’ Coding of a Training Set of Documents.” *Grossman and Cormack* at 26.

<sup>27</sup>*Moore v. Publicis Groupe*, 287 F.R.D. 182, 18 Wage & Hour Cas. 2d (BNA) 1479 (S.D. N.Y. 2012), adopted, 2012 WL 1446534 (S.D. N.Y. 2012).

<sup>28</sup>*Rio Tinto PLC v. Vale S.A.*, 2014 WL 7191250 (S.D. N.Y. 2014).

<sup>29</sup>*Hyles v. New York City*, 2016 WL 4077114 (S.D. N.Y. 2016) (“To be clear, the Court believes that for most cases today, TAR is the best and most efficient search tool . . . [t]he Court would have liked the City to use TAR in this case.”).

<sup>30</sup>*FCA US LLC v. Cummins, Inc.*, 2017 WL 2806896 (E.D. Mich. 2017).

<sup>31</sup>*Irish Bank Resolution Corp. v. Quinn* [2015] IEHC 175; *Pyrrho Investments Ltd. v. MWB Property Ltd.* [2016] EWHC 256 (Ch) (16 February 2016).

<sup>32</sup>*Da Silva Moore*, 287 F.R.D. at 189.

early engagement with the documents and use of analytics can provide significant advantages.

Traditionally, attorneys have approached voluminous discovery by proposing search terms. But, especially in the early stages of a case, “the way lawyers choose keywords is the equivalent of the child’s game of ‘Go Fish.’ The requesting party guesses what keywords might produce evidence to support its case without having much, if any knowledge of the responding party’s ‘cards’ (*i.e.*, the terminology used by the responding party’s custodians). Indeed, the responding party’s counsel often does not know what is in its own client’s ‘cards.’”<sup>33</sup>

Certain information can be gathered through client interviews: a basic understanding of what the client thinks the case is about, the kinds of documents that exist, holes where they know documents were lost or not preserved. But, the documents themselves may tell a different, or an additional, story. Reviewing document counts by custodian or timeframe may reveal holes where one would expect a large volume of documents, suggesting that data has been lost or some other issue has occurred. Reviewing a report of the types of documents in the document collection may identify a specific type of document one was not expecting that triggers questions: *e.g.*, does the client have a large number of x-ray files from a specific timeframe when counsel is not aware of any reason for such x-rays to exist? Reviewing the file structure of a server or shared drive may reveal collections of documents on an issue the client has not thought to flag. And, reviewing a computer-generated list of top concepts from the document collection may suggest that, while the client thinks the case is about sprockets, the parties were far more focused on widgets during a critical time period, which may warrant additional questions or investigation. All of this information can inform both the claim development process and the focus of, and strategy for, discovery.

This is also an important time to gather information regarding the potential burdens associated with discovery, preparing one to effectively advocate for a sensible, proportionate scope of discovery. Basic information, such as document volume and an estimated cost to review those documents, is absolutely necessary. But, it should only be viewed as a start; a great deal of additional information that can be gleaned.

If search terms have been proposed, reviewing a sample of records hitting on those terms can provide information about whether they are identifying a significant number of responsive documents. More detailed analysis of the responsiveness rate of

---

<sup>33</sup>Moore v. Publicis Groupe, 287 F.R.D. 182, 191, 18 Wage & Hour Cas. 2d (BNA) 1479 (S.D. N.Y. 2012), adopted, 2012 WL 1446534 (S.D. N.Y. 2012).

search terms by custodian, date, or other metadata category may suggest potential limitations that can be applied to particular terms to reduce the review population: for example, if a particular term has a high responsiveness rate in one time period, and a low responsiveness rate in another time period, a date limitation on the term may be appropriate.

A random sample of documents and analysis of the computer-generated concepts common to responsive materials may identify potential search terms that have not occurred to any party. Such a review also can uncover other trends that can inform the scope of discovery. Perhaps virtually all of a particular custodian's documents reviewed in a random sample are responsive, suggesting that that custodian should be considered presumptively responsive—or, virtually none of the custodian's documents are responsive, suggesting they should be excluded from review. Analysis of the data may also indicate, for instance, that all of the responsive emails for a particular high-volume custodian also include another custodian with a smaller number of documents or a higher percentage of responsiveness.

Taking the time to analyze and engage with the data provides early opportunities to better understand the case, and the scope of discovery provides crucial opportunities to take a peek at the cards in this high-stakes game of “Go Fish.”

### ***B. Document Review and Production***

As a case moves forward into discovery, litigants today face a choice: will they conduct a traditional, linear review of all documents, use TAR to inform and prioritize review, or review only a small subset of documents as part of a predictive coding approach? Each approach has benefits and drawbacks that must be weighed in the context of a particular case.

A traditional linear review, where each record is reviewed, organized by date, keyword relevance, or some other, human-defined factor, ensures that an attorney has eyes on every document. While attorney review isn't perfect or even necessarily better than TAR—studies suggest that computer systems can produce comparable results to human review<sup>34</sup>—a traditional linear review provides comfort that every document has been reviewed and considered, and no technological quirks or errors have led to unexpected results. Human reviewers can apply detailed coding to documents, allowing a party to engage in case analysis and preparation at the same time it works towards meeting its discovery obligations, while performing a detailed and nuanced assessment of privilege within the document set. However,

---

<sup>34</sup>See, e.g., *Da Silva Moore*, 287 F.R.D. at 190.

such a brute force approach can require significant time and money, and does not take advantage of the efficiencies that TAR can provide. And, where document volumes are large and deadlines are tight, this approach simply may not be feasible.

At the other extreme, predictive coding allows a party to code a small subset of documents to “train” a machine learning algorithm; the algorithm then propagates responsiveness coding to the rest of the documents. This approach allows a party to produce a large volume of documents in a short time, with comparatively-minimal expense. This can be an attractive approach, assuming a Rule 502 Order and/or strong claw-back agreement is in place to protect against the inadvertent production of privileged information. A party taking this approach, however, runs the risk of sending a production out the door without knowing what it includes: is there critical, case-changing information that hasn’t been identified? Are there litigation strategy documents that weren’t caught in privilege searches that, while they can be clawed back, will give the opposing party insight that provides a competitive advantage? Are there critical documents that are missed by the algorithm and not produced because they don’t have sufficient content to be analyzed? And, while this process allows for a comparatively inexpensive document production, it cannot address other case needs: once the documents have been produced, at least some portion of the documents almost certainly will need to be reviewed in the case-development process. A party also may elect to study the documents identified by the predictive coding algorithm prior to production, which alleviates some of these concerns, but which increases the time to complete production.

These first-generation, predictive coding methods are being replaced rapidly by Continuous Active Learning (“CAL”) technologies. CAL typically involves human study of a random sample of records. Once these documents are reviewed, the CAL algorithm scores the remaining materials based on the likelihood that those records are responsive; reviewers can then prioritize those documents with high responsiveness scores for study. As review continues, the CAL algorithm continues to “learn” and update the scores of remaining documents. This process bridges the gap between the two previous approaches: attorneys still consider the documents that are produced, but the most-likely-to-be-relevant materials are analyzed and integrated into the developing case first. As the responsiveness rate declines, sampling of the remaining documents may indicate that further review is unnecessary. But, as with predictive coding, records may elude the algorithm and not be produced.

Determining the best approach requires balancing multiple

factors: are there aggressive discovery deadlines; a prohibitively-large collection of potentially-responsive documents; significant budget constraints; complicated privilege issues that require careful attorney review; a large volume of documents that are not compatible with TAR; or pressing claim-development needs? Not every approach works for every case; understanding your documents, your case needs, and any additional pressures or constraints are all important in identifying the best approach or combination of approaches.

### ***C. Review of Documents Received in Discovery***

Once documents are received, a litigant has two primary challenges: find the important documents; and find any holes in the production.

Documents received in discovery can, like one's own records, be studied in various ways. A brute force, document-by-document review can ensure that no critical information is missed—especially those unexpected materials that one does not know to look for, but that can make a case. But, this sort of review, again, can be costly and inefficient—particularly in a construction case, where, for instance, the parties have exchanged voluminous, formal correspondence that already has been reviewed as part of the production review.

TAR and analytic techniques can be used in the same way they were used with a party's own documents. Study of an initial random sample of records can provide a broad overview of the kinds of materials in the collection. Analysis of various metadata, including file types and folder structure, may reveal information about what the producing party thought was important. Review of the common concepts in the documents may similarly confirm one's theories of what is important, or reveal unexpected topics that change the course of a case. All of these factors can inform one's priorities for review.

The same techniques and technologies used during review of one's own documents also can be useful. Allowing materials to be clustered based on their concepts, rather than organizing them in a purely linear fashion, may allow a single reviewer to see all of the pieces of a particular issue, rather than just a single moment in time, to put the pieces together more quickly. Using CAL or other predictive coding technologies to prioritize documents, based on review of prior records, can bring the important materials to the forefront, allowing one to analyze them much sooner.

Understanding what is *missing* from a production is often more challenging than understanding what is there. Analytics and review of metadata can be used in the same ways they were used with a party's own documents. Review of the metadata may reveal custodians or time periods for which records appear to be

missing. If a list of common concepts does not include critical issues that should exist, one may find, upon further review, that follow-up with the producing party—or an additional document request—is necessary.

There will come a point, in every complex construction case, where attorneys will need to expend a significant number of hours looking at the documents generated during a construction project and putting together a story. But, they need not go in blind; understanding and making use of the continually-evolving and—improving universe of electronic discovery technologies can help identify what is really important and reduce the time and cost it takes to get there.

#### ***D. Limitations of Technology***

TAR and other analytical methods can save significant time and costs, but for some data common in complex construction cases, over-reliance on technology is detrimental. This is especially so because one of the benefits of TAR is avoiding review of every document by focusing in on relevant ones; to take full advantage of these technologies, one must also understand, and account for, their limitations.

Consider, for instance, a box full of timesheets. A large construction project may generate dozens of such boxes. These records may prove vital for expert analysis, but may not warrant detailed document-by-document attorney review during the initial phases of discovery. These materials could be digitized and loaded into a database for review. However, timesheets are unlikely to provide sufficient—and sufficiently accurate—OCR text for TAR techniques to be effective, and they may not turn up in key-word searches, so predictive coding or CAL technologies may not find the relevant content. Document review attorneys will need to review each individual document pursuant to the review protocol. It likely will prove more efficient to conduct an old-fashioned, hard-copy review for responsiveness and privilege, log documents by box number or other identifying information so that they can be more easily located later, and then digitize relevant documents for production.

It similarly may be worthwhile to request that opposing parties make their paper files available for review. In addition to quickly identifying collections of documents, such as timesheets, critical documents, such as project diaries (which often are difficult to locate using search terms) can be identified and reviewed immediately, or logged for easy identification, once they are digitized.

Certain electronic materials common on a construction project likewise are poor candidates for current TAR techniques. Any record without text (e.g., photographs and diagrams), minimal

text (e.g., schedules and construction drawings), or with text in a non-narrative structure (e.g., spreadsheets and charts) is unlikely to be identified through TAR technologies that rely on the content of a document. To the extent that TAR is used to reduce document populations for review—whether for production or in review of opposing parties’ productions—these records should be identified by file extension, name, and other available metadata and segregated for separate review outside of the TAR process.

#### **IV. LITIGATION-READINESS PLANNING**

Having a litigation-readiness plan is much like having an on-site safety plan: waiting to draft a plan until an incident occurs just produces avoidable pressures and extra costs. With the appropriate plan in place, however, one can respond quickly and effectively.

Litigation-readiness generally refers to a series of policies, plans, and processes that allow a company to identify, preserve, collect and produce relevant documents and data in the event of litigation. In addition to company-wide litigation readiness, making sure that an individual construction project is litigation-ready will enable a rapid response when claims arise. This is particularly important because the more information that is available at the beginning of a case, the more effectively litigators can prepare for discovery and deploy TAR and analytics to more effectively and efficiently complete the discovery process and identify and analyze the critical case documents.

An effective litigation-readiness program includes the following steps:

- Identify a litigation-readiness team, including corporate IT, business and legal representatives;
- Establish, document, and enforce retention policies, taking into account litigation and regulatory retention obligations;
- Establish, document, and enforce BYOD policies for employees permitted to conduct company business on their personal devices;
- Develop a data map identifying and describing all potential sources of data;
- Evaluate software for data preservation and collection concerns;
- Establish and document a litigation hold procedure;
- Establish and document a collection procedure;
- Regularly review policies and procedures; and,
- Establish a process for releasing litigation holds.

On an individual project, the following steps can help ensure effective claim preparation:

- Establish and enforce document-management procedures.

Ensuring that records are kept in an orderly fashion (and one that can be explained to counsel) means that critical case materials can be located quickly, that complete collections of documents, such as schedules, can be assembled easily, and that key photographs can be located and connected to project issues. This preparation is particularly important for records that are not searchable by content, are not easily identifiable as stand-alone documents, or are not accessible without specialized software.

- Establish and enforce document-retention procedures. Ensuring that employees save key documents is particularly important as projects come to an end: it is no longer as simple as packing up the paper in the site trailer and sending it to storage, or even archiving employee email and file servers. If key information exists only on mobile devices, e.g., photographs or text messages, that data must be collected timely and preserved accessibly.
- Establish and enforce project communication procedures. If business is done by text message or through other messaging applications, then such data easily may be lost well before it is needed to prove an element of a claim or defense. Documenting agreements, admissions, and other critical information via a more permanent and identifiable method, such as letter or email, ensures that the information is available if needed.

## V. CONCLUSION

Meeting discovery obligations in today's project is a daunting task. Acting efficiently and effectively, while also advancing case development, requires both construction litigators and their clients to engage well before claims arise. Litigators must understand both the law and the benefits and limitations of available eDiscovery technologies. For their part, clients must implement litigation-readiness plans well in advance of any claims. While this will not, and cannot, wholly eliminate the substantial burden of discovery in major litigation, putting these pieces together provides the tools necessary to build a sound litigation strategy.