American Bar Association
Forum on the Construction Industry

Preservation, Collection, and Production of Documents and ESI in Construction Cases

John Foust
Jones Day
San Francisco, CA

and

Judge Timothy S. Hillman
United States District Court
District of Massachusetts

Presented at the 2012 Fall Meeting

Construction Counseling: Pulling Together for a Winning Strategy

October 18 – 19, 2012
Sheraton Boston Hotel, Boston, MA

I. The Challenges of Documents and ESI in Construction Cases

Document discovery and in particular electronic discovery can be one of the most challenging and costly aspects of litigation in the early twenty-first century. It also presents huge risks for litigants who, knowingly or not, fail to take the proper steps to preserve, collect, and produce relevant documents and electronically stored information ("ESI"). In order to avoid those risks, which can include the risk of receiving a terminating sanction or an adverse-inference jury instruction, it is imperative for counsel to stay aware of developments in the law of discovery and e-discovery and to work with their clients to ensure good hygiene with respect to the preservation, collection, and production of documents.

The field of construction law presents major challenges for a litigant seeking to practice good document hygiene. First, construction projects run on documents. On large commercial projects, the official hard copy job files can easily range from 2,000 to 15,000 banker boxes. Add to that electronic files such as emails, server files, accounting databases, change order databases, design document depositories, and countless files saved on personal devises like laptops, desktops, external hard drives, and even iPhones and BlackBerries. The result is an enormous universe of potentially discoverable documents.

A second way in which construction law presents a challenge with respect to document discovery is that the scope of relevance in most construction disputes is frequently very broad and often extends to all documents that relate to the construction project. Unlike other kinds of disputes that are limited to specific items or events, like patent disputes or fraud claims, construction disputes—once they are fleshed out with counterclaims and cross-claims against subcontractors, designers, construction managers, and even lenders—often implicate every aspect of the project, from conceptual design to operation of the completed project. This makes

it difficult to cull relevant documents and generally results in parties producing everything from the project.

A third challenge for document discovery in construction cases is that the tools and generally accepted best practices in the field of electronic discovery are not particularly well-suited for construction disputes. Predictive coding and keyword searches, which may be useful in patent cases, securities cases, antitrust cases, and employment cases, provide limited value in most construction cases. Moreover, the products and pricing structures peddled by e-discovery vendors are really designed for email and other text-readable data. Using those products in construction cases with CAD files, Primavera schedule files, shop drawings, handwritten field notes, and construction photos can be frustrating and expensive. None of those file types can be searched using keywords, and most are gigabyte heavy and therefore expensive to process and host under the typical per-gigabyte vendor pricing structure.

This paper identifies the challenges of document discovery in the context of construction disputes and then suggests some practices for dealing with those challenges in a way that is practical and cost efficient. The first section focuses on the duty to preserve documents, including ESI, that attaches once litigation is anticipated. The second section discusses the collection and production of ESI once litigation has been filed. Each section begins with a summary of the current state of the law and then offers suggested practices for interpreting and applying that law in the context of a construction dispute.

II. Preservation of Documents, Including ESI

A. *The Current State of the Law Regarding the Duty to Preserve*

Every lawyer is familiar with the notion that at some point in the development of a dispute the parties have a duty to take active steps to preserve documents—including both hard

copy documents and electronically stored information or "ESI" —that may be relevant to the issues underlying the dispute. The duty to preserve documents, including both hard copy documents and ESI, attaches as soon as there is "anticipation of litigation."[1] When a party is not anticipating litigation, there is no duty to preserve documents unless otherwise indicated or required by state or federal statute, judicial decision, or regulations governing the ordinary course of business.

One anticipates litigation when a party is placed on notice of a credible probability that one will become involved in litigation, or when a party seriously contemplates initiating litigation, or when a party takes specific actions to commence litigation.[2] In considering whether litigation was anticipated, courts have looked at sheer magnitude of loss, a plaintiff's attempt to document damages, and a plaintiff's retention of experts and counsel.[3] Other courts have ascertained that a duty to preserve can arise well before litigation as a result of discussion of plans to bring suit,[4] when one thinks they are getting sued,[5] when a party has knowledge of a potential claim,[6] and at the latest when a party is in receipt of a complaint.[7] Typically, plaintiffs have been found to necessarily anticipate litigation at least as of the time the complaint is filed.[8]

The scope of the duty to preserve documents extends to all documents that are relevant to the subject of the litigation or anticipated litigation and that are within a party's possession, custody, or control.[9] In deciding the relevancy of documents, courts have held that there is a duty to preserve all evidence having a tendency to make existence of any fact that is of consequence to the determination of the action probable or less probable than it would be without the evidence.[10] A court may limit the scope of preservation of electronically stored information when it is unreasonably cumulative, duplicative, available elsewhere, inaccessible, or when the burden or expense of preservation outweighs its likely benefit.[11]

When a duty to preserve arises, it imposes several affirmative obligations on parties and their counsel. First, parties and their counsel have a duty to issue a litigation hold and to suspend all routine document retention/destruction policies.[12] Second, counsel must oversee compliance of document retention and production by parties and key players.[13] This includes locating all sources of potentially relevant data, both active and stored, and becoming fully aware of all document retention policies and data retention architecture by talking to IT personnel.[14] It also includes communicating with all key players in the litigation and educating them about their preservation obligations.[15]

B. *Interpreting and Applying the Duty to Preserve in the Context of the Construction Industry*

Applying the law regarding a party's duty to preserve documents in the context of the construction industry can be challenging. First, there is the matter of determining the point at which litigation is anticipated and the duty to preserve arises. On large construction projects, the parties are constantly negotiating change orders and other interim payment disputes. Sometimes the parties are able to reach an agreement on these disputed items and a change order is signed or a pay application is approved. Other times, however, the parties defer the disputed items and attempt to resolve everything through negotiations at the end of the job. Determining whether there is an anticipation of litigation under these circumstances will likely turn on specific facts such whether a formal written claim was made, whether a claims consultant was retained, whether litigation was threatened or hinted at during negotiation, and whether a party put its carrier on notice.

Once it is determined that there is a duty to preserve documents in a construction dispute, counsel should take certain steps to ensure that the necessary documents, including hard copy

documents and ESI, are in fact preserved.[16] At the very least, this should include working with the client to issue a litigation hold.[17] It also should include talking to the correct people within the client organization, such as the head of information technology (IT) and the records manager, and instructing those persons about suspending all routine document destruction like automatic email deletion.[18]

In addition to personnel from records and IT, counsel should reach out to the heads of relevant departments such as accounting and internal audit regarding the preservation of hard copy records and ESI. Counsel should also determine whether there are active databases, including ones hosted by third-party vendors, that need to be preserved possibly by creating snapshots or mirror copies. Finally, because ESI is dynamic and subject to change notwithstanding reasonable preservation efforts, the safest course is to take a snapshot or make a backup copy of all potentially relevant data at the time the duty to preserve arises. This is particularly true for custodian data—roughly defined to include ESI associated with and controlled by a single custodian such as local drives, user network drives, and emails—since custodians are notoriously unreliable in adhering to litigation hold directives.

III. Collection and Production of Documents, Including ESI

A. *Current State of the Law Regarding the Collection and Production of Documents*

1. *Hard-Copy Documents*

There has not been much recent development in the law surrounding the collection and production of hard-copy documents. The general rule is that all relevant, responsive, non-privileged materials and documents that are within a party's possession, custody, or control must be produced.[19] The documents must be produced "as they are kept in the usual course of

business" or else must be organized and labeled to correspond to the specific categories in the document requests."[20]

There is a growing trend for courts to set protocols regarding the format of the production of hard-copy documents. One of the more common requirements is to require that parties either (A) produce the originals for inspection or copying in the same organizational structure in which they were maintained in the ordinary course of business or (B) produce scanned images that have been processed using optical character recognition (OCR) software and packaged into a load file that may or may not be required to contain metadata fields such as the location from which the documents were collected and the Bates number.

2. *Electronically Stored Information (ESI)*

In the quickly-developing field of electronic discovery, the standard accepted practice as articulated by several high-profile cases from the federal courts is that parties are obligated to collect and produce ESI belonging to certain "key players." This term "key players" is taken from the *Zubulake* cases, which were among the first to explore the new frontier of ESI preservation and production.[21] The idea is that because of the potentially overwhelming volume and cumulative nature of ESI, parties are not typically expected to collect, process, review, and produce every single file of potentially relevant ESI. Instead, parties start by identifying the "key players" who acted as the custodians of relevant information.[22] Parties should then collect and produce all relevant, responsive, and non-privileged ESI created or maintained by the key players.[23]

When producing ESI, "a party must produce it in a form or forms in which it is ordinarily maintained or in a reasonably usable form or forms."[24] As a matter of practice, ESI is typically produced in either its native format or as image files (PDF or TIFF).[25] There are advantages and

disadvantages associated with native productions and image files as discussed below in the section on recommended practices in construction disputes.

Parties must take into account the need to collect and produce reasonably relevant metadata.[26] Metadata is information associated with an electronic file that is not necessarily viewable when the file is opened in its intended application. Metadata can provide information about the file such as when it was created, who it was created by, when it was last modified, where it was stored, and other historical information.[27] Parties must take care to preserve the metadata during the collection of ESI, since copying the files without using the correct software can alter metadata fields such as the created date. Parties must also produce ESI with the metadata intact. Courts have stated that metadata is an integral and intrinsic part of any electronic record.[28]

Courts do not typically require the collection or production of ESI that is not reasonably accessible. Such a determination is dependent on the circumstances and anticipated relevancy, but typically a party is not required to collect or preserve ESI that it does preserve in the ordinary course of business, such as deleted, slack, fragmented, unallocated, or ephemeral data.[29] Similarly, unless the opposing party can make a specific showing of relevancy or need, courts have typically allowed parties to avoid unduly burdensome, costly or duplicative ESI productions. For example, typically "[a] party need not produce the same electronically stored information in more than one form."[30] Moreover, Rule 26 of the Federal Rules of Civil Procedure provides a specific limitation on the production of ESI that is not "reasonable accessible because of undue burden or cost."[31] Thus, parties can generally avoid costly restores of tapes, silos, and other long term retention media unless and until a showing a heightened relevancy of that data is shown.

Because of the need to tailor e-discovery to the facts of a particular case, it is becoming increasingly common for courts to enter orders establishing the protocol for the collection and production of ESI. This is particularly true in large, multi-party cases. It is also becoming common in large cases for courts to appoint a special master to oversee electronic discovery, ideally someone with technical knowledge who can help craft and oversee compliance with the protocol for collecting and producing ESI.

B. *Recommended Practice for Collecting and Producing Hard-Copy Documents and ESI in the Context of a Construction Dispute*

1. *Producing Hard-Copy Documents in Construction Cases*

Counsel should begin the task of collecting and producing hard copy documents by identifying and locating the record set of documents maintained by project personnel. The record set of documents varies from project to project but typically includes the usual construction documents such as change documentation, schedules, drawings, correspondence, daily/weekly/monthly work reports. Counsel should try to identify all potential locations for these records through desk-side interviews with the people tasked as the custodians of the official records, such as the client's records manager, controller, and even the project manager. Locations to look out for include off-site warehouses, the project office, the home office, job site trailers, and storage rooms. Ideally, the client will be able to provide indices or other documentation regarding the organizational structure of the documents.

In addition to what was just described as the record set of project documents, counsel will also need to identify and locate what might be described as the working files. The working files often include the same types of documents that exist in the record set, but they are maintained and used by the individuals involved with the project, including those who might be described as

key players. They are often kept in the personal offices of the potential witnesses, but are sometimes squirreled away in desks, drawers, briefcases, and home offices. Counsel should start by talking to the key players from all relevant departments—such as accounting, project controls, internal audit, operations, etc.—to determine the extent and locations of such working files.

As for the format of the production, counsel has two choices—either make the originals available for inspection and copying, which the court may require to be accompanied by an index,[32] or produce scanned images, which the court may require to be processed using OCR technology.[33] Logistically, it is almost always preferable to produce working copies as scanned electronic images. That way, the client can continue its day-to-day operations (in addition to assisting with the litigation) with minimal interruption, and counsel will have a manageable electronic record of the documents it needs to review and produce. Counsel will often decide, however, that the record copies—particularly the voluminous warehouse files and project archives maintained by general contractors and owners—are more efficiently produced by making the originals available for inspection and copying, which is undeniably less expensive.[34]

2. *Producing ESI in Construction Cases*

This paper began with a discussion of the challenges that a construction lawyer faces in connection with the preservation, collection, and production of ESI. The vast amount of data, the types of data, and the large scope of relevance makes collecting and producing ESI a daunting and potentially cost-prohibitive task. However, with some thoughtful planning at the front end and careful execution throughout, it is possible to satisfy the obligation to produce all relevant, responsive, non-privileged ESI in a reasonably cost-efficient manner.

The key to handling the collection and production of ESI in a defensible and cost-efficient manner is to begin by identifying the various sources and types of relevant data and then

building a plan that is tailored for each data source and data type. In other words, different data sources and types have different characteristics that may require different treatment. For example CAD files, emails, and a cost-accounting database have nothing in common except that they all involve digital files. The first involves large files that contain no text-readable data and cannot be viewed without specialty software. The second are commonplace files that are text readable and searchable according to useful metadata fields like, date, time, sender, recipient, and subject. The third is a database filled with numbers that are useless without the functionality and reporting capabilities that are inherent to the database. It makes no sense to treat these three data sources the same.

The starting point is to locate all of the potential sources of ESI. Presumably this was already done in connection with the preservation efforts discussed in the previous section of this paper. It involves interviewing records custodians, IT personnel, and department heads to determine how and where data is stored. Likely sources are: email files located on the company server; email files located on custodian local drives; other custodian local files; custodian network data located on the company server; shared drives or department directories located on the company server; databases and document management systems hosted by the client; databases and document management systems hosted by third-party vendors; and loose storage media such as CDs, DVDs, and flash memory that is often found scattered among the hard-copy files.

The types of data located will vary from project to project depending on the software and the systems used in connection with the project. Generally speaking, however, the data can be classified into the following types for the purpose of managing the document collection and production.

(a) Text-readable data. These are files like emails, Word documents, PowerPoint slides, Excel spreadsheets, and some PDFs and TIFFs that contain a layer of text. These file types may benefit from the tools offered by most e-discovery vendors like early case assessment, predictive coding, and keyword searches.[35]

(b) Non-text-readable data. These are files that do not contain a readable text layer. Construction projects often involve large stores of non-text-readable data, like CAD files, schedule files (Primavera and MS Project), and job site photos. It is important to process and manage these files separately because they will not benefit from the searchability offered with expensive e-discovery tools. Moreover, many of these file types are only viewable with specialty software that attorneys are not likely to have.[36]

(c) Scanned PDFs and TIFFs. PDFs and TIFFs are image files, which means that they are essentially photographs rather than an organization of text or numbers. What makes these types of PDFs and TIFFs tricky from an e-discovery point of view is that sometimes they are text-readable, sometimes they are *non*-text-readable, and sometimes they are non-text-readable but are capable of being made text-readable with the use of OCR technology.[37] The ones that are text-readable should be treated the same as the rest of the text-readable data, and counsel should seriously consider having the non-text-readable data OCR processed so that it too can be treated as text-readable. The remaining non-text-readable files, which are likely to be made up of either handwritten notes or graphical images, will need to be set aside for a page-by-page review unless counsel can make a ruling on responsiveness, relevance, and privilege based on the source of the files.

(d) Hosted data. Construction projects often rely on outside vendors to host project data, including design-related documents (*e.g.*, Buzzsaw, Planwell, etc.) and

change-management documents (*e.g.*, Skire, Aconex, etc.). Hosted data will need to be produced in a way that preserves the functionality associated with the platform.

(e) Databases. Most construction projects use cost accounting databases (*e.g.*, CGC, Timberline); many use a change management database (*e.g.*, Prolog); and some use databases to log other project-related information. The challenges with databases is that the functionality is tied up in the orientation of the database and often will not survive if the data is dumped into an export file and processed. Accordingly, it is important to treat databases differently than other ESI and to cooperate by providing insight into the database structure and reporting characteristics as discussed further below.

Once counsel has determined the location of all relevant ESI and has determined what file types are expected to exist among the various sources, the next step is to devise a plan for handling each data source and type in a practical and cost efficient manner.

- Segregate text-readable files. It is good practice to identify and segregate all of the files that contain text-readable data, like Word files, Excel files, emails (PSTs, MSGs), and PDFs with an OCR text layer, and then have those files processed and uploaded into an e-discovery platform for searching and tagging.[38] This will allow counsel to: (1) apply inclusionary and/or exclusionary terms to filter the data for responsiveness; (2) make use of technology such as predictive coding or early case assessment tools; (3) pay a premium to have this data processed and hosted to ensure maximum functionality; and (4) screen the data for privilege and perform substantive document review using keywords and critical metadata fields.

- Process non-text-readable data efficiently. Rather than sending all non-text-readable data to an e-discovery vendor for expensive processing and hosting on a searchable platform, it is important to consider cost saving alternatives that would provide the

same level of functionality at a greatly reduced cost. Non-text-readable files are often very large files (e.g., CAD files and Primavera files) and therefore expensive to process under traditional fee arrangements, yet these files derive minimal benefit from the processing and hosting offered by ESI vendors because they cannot be searched using keywords. Thus, it is most practical and cost-effective to produce non-text-readable data as native files without any processing. Or, if there is an order in place requiring all ESI to be processed, it is worth looking for a vendor that will perform the minimum processing permitted by the ESI protocol on this non-text-readable data for an hourly rate rather than a per-gigabyte fee structure.[39]

- Consider producing native files. There is some disagreement on this subject in the e-discovery world, but counsel should at least consider the prospect of negotiating an ESI protocol that requires the production of native files.[40] Native production provides the most functionality and transparency for the documents, and theoretically a cost savings from foregoing conversion to PDF or TIFF. However, many vendors aren't equipped to deal with native productions as their review platforms are based on PDF and TIFF formats, and thus some vendors will converting documents for use in their review platform anyway. Another challenge that arises in native productions is Bates numbering. Native files cannot be Bates numbered in a manner that maintains the integrity of the original file (*i.e.*, it is not practical or forensically sound to add a Bates number to the footer of a native Excel file). At the very least, parties should consider protocols for separately producing Excel, Access, Primavera, CAD, and other file types where the functionality of the file goes beyond reading the text or looking at a picture.

- Provide access to databases and hosted data. As mentioned above, many construction projects involve databases and hosted data. For accounting and other

databases that often require special software licenses, counsel should consider exchanging written lists of reporting capabilities and agreeing to run reports on an as-requested basis. For change-management databases, counsel should consider making the database available via remote server or a full output file. For hosted data, counsel should consider providing user IDs and login information to the parties, which may require vendor approval.

- Use digital parameters to help structure document review. When reviewing ESI for privilege or substance, counsel will have the ability to structure the review around the following key parameters so long as the data was processed correctly: (1) custodian/source; (2) file type; (3) dates from metadata; (4) keywords where applicable; (5) email sender/recipient where applicable; and (6) network location/folder path. If this is done in a thoughtful way, it will drastically reduce the amount of data that has to be individually reviewed by a reviewer.

IV. Parting Thoughts

Document discovery and in particular electronic discovery is expensive, time consuming, and fraught with challenges. That is particularly true in the context of large construction disputes where the volume of documents is high, the scope of relevance is broad, and the traditional e-discovery tools are not well designed to deal with many of the sources and file types common in the industry. Therefore, construction lawyers need to take care at the outset to conduct a thorough investigation to identify all potential sources and types of documents and ESI and then devise a plan for collecting and producing those documents in a way that treats each type and source according to its own specific needs. Given the costs and consequences of missteps, the issues described above require serious consideration by lawyers and clients alike. Hopefully,

the rules and practices discussed in this paper provide some helpful guidance on how to go about that task in a fulsome yet cost efficient manner.

---

[1] *Zubulake v. UBS Warburg, LLC*, 220 F.R.D. 212, 216 (S.D.N.Y. 2002) (hereinafter *Zubulake IV*).

[2] *Id.*

[3] *Pension Comm. of the Univ. of Montreal Pension Plan v. Banc of America Securities, LLC*, 685 F. Supp. 2d 456, 466 & n.27 (S.D.N.Y. 2010).

[4] *Rimkus Consulting Grp., Inc. v. Cammarata*, 688 F. Supp. 2d 598 (S.D. Tex. 2010).

[5] *Cenveo Corp. v. S. Graphic Sys., Inc.*, 2010 WL 3893680 (D. Minn. June 18, 2010).

[6] *Hogan v. Raymond Corp.*, 777 F. Supp. 2d 906 (W.D. Pa. 2011).

[7] *Winters v. Textron, Inc.*, 187 F.R.D. 518, 520 (M.D. Pa. 1999).

[8] *Cyntegra, Inc. v. Idexx Laboratories, Inc.*, 2007 WL 5193736 (C.D. Cal. Sept. 21, 2007).

[9] Thomas Y. Allman, *E-Discovery in Federal and State Courts: The Impact of the 2006 Federal Amendments*, June 14,2012, http://blog.hinshawlaw.com/practicalediscovery/wp-content/uploads/2012/06/2012FedStateEDiscoveryRulesJune14.pdf; *Wm. T. Thompson Co. v. Gen. Nutrition Corp.*, 593 F. Supp. 1443, 1455 (C.D. Cal. 1984). The duty to preserve documents also extends to all documents reasonably calculated to lead to discovery of admissible evidence, reasonably likely to be requested during discovery, or that are the subject of a pending discovery request. *Id.*

[10] *Zubulake v. UBS Warburg, LLC*, 229 F.R.D. 422 (S.D.N.Y. 2004) ("*Zubulake V*").

[11] Allman, *supra* note 9, 43. Courts weigh the burden of preservation against certain factors to determine whether the duty to preserve would place an undue burden on a party.

Pippins v. KPMG, LLC, 2011 WL 4701849 (S.D.N.Y. Oct. 7, 2011). In particular, courts consider potential relevance of information, whether the information related to key players, extent to which the information was duplicative or cumulative, needs of the case, amount in controversy, resources available, the importance of the issues, and the importance of discovery in resolving those issues. *Id.* Moreover, inaccessibility for production does not necessarily excuse preservation. Allman, *supra* note 9, at 21. Temporary cache files, RAM, or wave files, which are consistently overwritten, may need to be preserved in case there is a specific request for them. *Id.*

[12] *Zubulake V*, 229 F.R.D. at 431, citing *Zubulake IV*, 220 F.R.D. at 219. The litigation hold does not necessarily apply to back-up tapes maintained for disaster recovery, but it does apply to those that are accessible and actively used for information retrieval. *Id.*

[13] *Zubulake V*, 229 F.R.D. at 432.

[14] *Id.*

[15] *Id.*

[16] For a set us useful guidelines see generally: The Sedona Conference, Commentary on Preservation, Management and Identification of Sources of Information that Are Not Reasonably Accessible, July 2008 ("Sedona Commentary on Preservation").

[17] *See Pension Comm. of the Univ. of Montreal*,685 F. Supp. 2d at 466 (holding that failure to issue a litigation hold may be gross negligence); *see also generally* The Sedona Conference, Commentary on Legal Holds, September 2010.

[18] *See* Sedona Commentary on Preservation, *supra* at note 16, at 8.

[19] Fed. R. Civ. Proc., Rule 26(b): "Discovery Scope and Limits. (1) Scope in General. Unless otherwise limited by court order, the scope of discovery is as follows: Parties may obtain

discovery regarding any non privileged matter that is relevant to any party's claim or defense. … Relevant information need not be admissible at the trial if the discovery appears reasonably calculated to lead to the discovery of admissible evidence."

[20] Fed. R. Civ. Proc., Rule 34(b)(2)(E)(i).

[21] *See generally Zubulake IV*, 220 F.R.D. at 218; *Zubulake V*, 229 F.R.D. at 432; s*ee also generally Pension Comm. of the Univ. of Montreal*, 685 F. Supp. 2d at 465 (failing to collect records, including hard copy and ESI, from key players constitutes gross negligence or willful breach of a party's duty to preserve and produce).

[22] The court in *Zubulake IV* defined key players as "those employees likely to have relevant information." *Zubulake IV*, 220 F.R.D. at 218 (emphasis added). Later cases adopted *Zubulake IV*'s ruling with respect to key players and emphasized that it is the quality or nature of an individual's involvement with certain issues or events which determines that individual's position as a key player—the amount of interaction and length of involvement are not determinative factors. *See Goodman v. Praxair Services, Inc.*, 632 F. Supp. 2d 494, 512 (D. Md. 2009) (discussing volume of interaction); *Huggins v. Prince George's County, Maryland*, 750 F. Supp. 2d 549, 561 (D. Md. 2010) (discussing length of involvement). In *Goodman*, even a non-supervisory employee with limited contact with the plaintiff and "somewhat limited involvement" in the project-at-issue was considered to be a key player because he had discussed the project and it was possible he had emails that "could be relevant." *Goodman*, 632 F. Supp. 2d at 516. In *Huggins*, an employee who was not involved with a property dispute from the start was considered to be a key player despite being involved in only the aftermath. *Huggins*, 750 F. Supp. 2d at 561. Both cases re-emphasize the principle from *Zubulake IV* that key players are "employees likely to have relevant information." *Zubulake IV*, 220 F.R.D. at 218.

[23] Allman, *supra* note 9, at 27.

[24] Fed. R. Civ. Proc., Rule 34(b)(2)(E)(ii); *see also Adams v. AllianceOne, Inc.*, 2011 WL 2066617 (S.D. Cal. May 25, 2011) (holding that where a plaintiff does not specify a format in his discovery request, a producing party is not required to produce ESI in the format it is ordinarily maintained so long as it is produced in a reasonably useable format).

[25] Allman, *supra* note 9, at 58.

[26] *Id.*

[27] *See* The Sedona Conference, The Sedona Conference Glossary, Second Edition: E-Discovery & Digital Information Management (2007) ("Sedona Glossary"). Metadata is defined in the Sedona Glossary as follows: "Data typically stored electronically that describes characteristics of ESI, found in different places in different forms. [Metadata] [c]an be supplied by applications, users or the file system. Metadata can describe how, when and by whom ESI was collected, created, accessed, modified and how it is formatted. [Metadata] [c]an be altered intentionally or inadvertently. Certain metadata can be extracted when native files are processed for litigation. Some metadata, such as file dates and sizes, can easily be seen by users; other metadata can be hidden or embedded and unavailable to computer users who are not technically adept. Metadata is generally not reproduced in full form when a document is printed to paper or electronic image."

[28] *Nat'l Day Laborer Organizing. Network v. U.S. Immigration & Customs Enforcement Agency*, 2011 WL 381625 (S.D.N.Y. Feb. 7, 2011) (Opinion and Order withdrawn by order from the court dated June 17, 2011 due to the resolution between the parties and incomplete record).

[29] *But see Columbia Pictures Indus. Inc. v. Bunnell*, 245 F.R.D. 443 (C.D. Cal. 2007) (Magistrate Judge Jacqueline Choolijan, following Ninth Circuit precedent, ordered defendants to preserve ephemeral data stored in RAM).

[30] Fed. R. Civ. Proc., Rule 34(b)(2)(E)(iii).

[31] Fed. R. Civ. Proc., Rule 26(b)(2)(B).

[32] Counsel should consider preparing an index even if not required to do so. One of the major downsides to producing originals is that it is not practical or cost-efficient to Bates-label large volumes of original documents. In other words, if you are incurring the cost to handle each individual page to apply a Bates stamp, you might as well incur the incremental cost of imaging the document and then produce the scanned images. The lack of Bates numbering can make it difficult to prove that a document was produced or trace it back to its original source for the purpose of laying a foundation at trial. An index that catalogs the documents by box or container number can help mitigate the traceability issues caused by the lack of Bates numbering.

[33] Even if the court does not require scanned documents to be OCR processed, it is generally good practice to do so. The increased searchability will assist counsel during the privilege screen as well as during the substantive review. When counsel chooses to perform OCR processing, even if not required to do so, most courts will require the documents to be produced with the OCR layer so that the opposing parties receive the benefit of the advanced functionality added by OCR.

[34] Issues of attorney-client privilege and the work-product doctrine are beyond the scope of this paper, but it is worth noting that counsel will need to consider the logistics of screening the originals for privilege before making them available for inspection and copying. This can be done in advance of the production, or for larger productions where a quality index has been

provided counsel may be able to get away with screening the boxes on a rolling basis as they are requested by the opposing parties.

[35] The following is a non-exhaustive list of text-readable file types: ace, arc, arj, arx, bh, cal, cat, csv, dat, db, dbx, doc, docx, dot, dotm, dotx, dst, dstx, efx, email, eml, gnu, gra, gz, gzip, jar, lha, mbox, mbx, mdb, mde, mde, mpp, msg, nsf, ost, pdf, pnf, pot, potx, ppt, pptx, pst, pub, rar, rpt, rtf, shw, tar, taz, thmx, tif, tiff, txt, tz, wbk, wk4, wmf, wpd, wps, wri, xla, xlam, xlm, xls, xlsm, xlsx, xlt, xltx, z, zip. Note that some of these files, like .zip, are container files that may or may not contain text-readable files. For purposed of collection and production, however, it may be necessary to treat container files like text files.

[36] The following is a non-exhaustive list of some non-text-readable file types: ac$, atc, avi, bmp, cdd, cdr, dsp, dtd, dwf, dwg, dws, dwt, dxf, gif, hdi, jpeg, jpg, mov, mp3, mp4, mpeg, mpg, p3, pdd, plt, png, prx, rfa, shx, vsd, vss, vst, wav, wpg.

[37] PDFs and TIFFs are image files, similar to photographs, that do not contain searchable text. However, sometimes the image file is accompanied by a hidden text layer. This typically arises when the image file was converted from another program such as Word or Excel or when it has been processed using OCR technology.

[38] It is important that family members (*i.e.*, emails and their attachments) are <u>not</u> separated from one another during this process. In other words, a non-text-readable document that is an attachment to a text-readable email should be kept with its parent email and processed with the rest of the email.

[39] As with all ESI, make sure to preserve all metadata and to capture all relevant information regarding the location, access, and name of these non-text-readable files during collection and processing.

[40] The most common practice in the industry is for documents to be converted into image files (typically single-page TIFFs or multi-page PDFs) and produced in that format. However, this is problematic for all document types *other than* emails and word-processing files. First, many documents were not formatted for printing and will therefore be converted into awkward image files that do not resemble the native version in any meaningful way. Second, important information such as formulas and hidden columns (in Excel) are left behind during the conversion.