

143 T.C. No. 9

UNITED STATES TAX COURT

DYNAMO HOLDINGS LIMITED PARTNERSHIP,
DYNAMO, GP, INC., TAX MATTERS PARTNER, Petitioner *v.*
COMMISSIONER OF INTERNAL REVENUE, Respondent

BEEKMAN VISTA, INC., Petitioner *v.*
COMMISSIONER OF INTERNAL REVENUE, Respondent

Docket Nos. 2685-11, 8393-12.

Filed September 17, 2014.

R requests that Ps produce electronically stored information contained on two backup storage tapes or, alternatively, the tapes themselves (or copies thereof). Ps acknowledge that the tapes contain tax-related information but assert that the tapes also contain privileged information that Ps have a right or duty to protect. Ps assert that they must review the responsive information on the tapes before giving the information to R to ensure that privileged or confidential information is not disclosed. Ps request that the Court let them use "predictive coding", a technique prevalent in the technology industry but not yet formally sanctioned by this Court, to help identify the information that is responsive to R's request.

Held: Ps may use predictive coding in responding to R's request.

SERVED SEP 17 2014

Martin R. Press, Edward A. Marod, Lu-Ann Mancini Dominguez, and Alan Stuart Lederman, for petitioners.

David B. Flassing and Lisa Goldberg, for respondent.

OPINION

BUCH, Judge: These consolidated cases are before the Court on respondent's motion to compel production of documents.¹ The cases concern various transfers from Beekman Vista, Inc. (Beekman), to a related entity, Dynamo Holdings Limited Partnership (Dynamo). Respondent determined that the transfers are disguised gifts to Dynamo's owners. Petitioners assert that the transfers are loans.

Respondent requests that petitioners produce the electronically stored information (ESI) contained on two specified backup storage tapes or, alternatively, that they produce the tapes themselves (or copies thereof). Petitioners assert that it will take many months and cost at least \$450,000 to fulfill respondent's request because they would need to review each document on the tapes to identify what is responsive and then withhold privileged or confidential

¹Respondent also moved to compel interrogatories. We will separately address that motion in an order.

information. Petitioners request that the Court deny respondent's motion as a "fishing expedition" in search of new issues that could be raised in these or other cases. Alternatively, petitioners request that the Court let them use predictive coding, a technique prevalent in the technological industry but not yet formally sanctioned by this Court, to efficiently and economically identify the nonprivileged information responsive to respondent's discovery request.

Respondent counters that he wants the backup tapes to review the ESI's metadata and verify the dates on which certain documents were created. Respondent states that he also wants the backup tapes to ascertain all transfers relevant to this proceeding. Respondent opposes petitioners' request to use predictive coding because, he states, predictive coding is an "unproven technology". Respondent adds that petitioners need not devote their claimed time or expense to this matter because they can simply give him access to all data on the two tapes and preserve the right (through a "clawback agreement") to later claim that some or all of the data is privileged information not subject to discovery.²

²We understand respondent's use of the term "clawback agreement" to mean that the disclosure of any privileged information on the tapes would not be a waiver of any privilege that would otherwise apply to that information.

The Court held an evidentiary hearing on respondent's motion. We will grant respondent's motion to the limited extent stated herein. Specifically, we hold that petitioners must respond to respondent's discovery request but that they may use predictive coding in doing so.

Background

I. Relevant Entities

A. Beekman

Beekman is a corporation wholly owned by a Canadian entity which is controlled by Delia Moog. Beekman's mailing address was in Florida when its petition was filed.

B. Dynamo

Dynamo is a limited partnership owned by a corporation and two trusts that were established for Ms. Moog's daughter and nephew. Dynamo's tax matters partner is Dynamo GP, Inc. Dynamo, through its tax matters partner, alleges that its principal place of business was in Delaware when its petition was filed.

Respondent alleges that Dynamo's principal place of business was in Florida at that time.

II. Backup Tapes

Dynamo backs up onto tapes its entire exchange server (inclusive of emails, operating system, and configuration information). Dynamo performs this backup work every four weeks and at the end of every month. Dynamo generally retains its backup tapes for one year.

Respondent seeks two of the backup tapes, specifically, the "Month End August 2010 ORANGE" and the "Month End Jan 08 ORANGE". These tapes contain data backed up from (1) an exchange server and (2) a domain controller and file server (KSH-DC). The exchange server database has approximately 200 mailboxes ranging in size from 500 megabytes to 1 gigabyte each. The KSH-DC has a common group and a user group. The common group has shares where assigned users may store data to be shared with other assigned users. The common group has approximately 50 common top-level file shares and an undetermined number of subfolders, and ownership of these files may not be limited to the authors of the documents. The user group is in a section of the network assigned to a specific individual and has approximately 200 user share folders.

III. Petitioners' Request To Use Predictive Coding

Petitioners acknowledge that the two requested backup tapes contain tax-related information but assert that the tapes also contain "personal identification information, health insurance information, HIPAA protected information and other confidential information that Petitioners have a duty to protect."³ Petitioners assert that if they must respond to respondent's discovery request, they must review the documents on the backup tapes to ensure that no privileged or confidential information is disclosed before giving any information to respondent. Petitioners ask the Court to let them use predictive coding to efficiently and economically help identify the nonprivileged information that is responsive to respondent's discovery request. More specifically, petitioners want to implement the following procedure to respond to the request:

1. Restore some or all of the data from the tapes.
2. Qualify the restored data; i.e., remove NIST files, system files, etc.^[4]

³The Health Insurance Portability and Accountability Act of 1996 (HIPAA), Pub. L. No. 104-191, secs. 261-264, 110 Stat. at 2021-2033, contains privacy rules and gave rise to privacy regulations relating to individually identifiable health information.

⁴The National Institute of Standards and Technology (NIST), which is an agency of the U.S. Department of Commerce, maintains a database of hash values
(continued...)

3. Index and load the qualified restored data into a review environment.

4. Apply criteria to the loaded data to remove duplicate messages and other nonrelevant information.

5. Through the implementation of predictive coding, review the remaining data using search criteria that the parties agree upon to ascertain, on the one hand, information that is relevant to the matter, and on the other hand, potentially relevant information that should be withheld as privileged or confidential information.

6. Produce the relevant nonprivileged information and a privilege log that sets forth the claimed privileged documents and sufficient information supporting that claim.

Discussion

I. Discovery in General

A party in this Court generally may obtain discovery of documents and ESI to the extent that the information contained therein is not privileged and is relevant to the subject matter of the case. See Rule 70(a)(1) and (b);⁵ see also Rule 72(a).⁶

⁴(...continued)
of files that typically are part of an operating system or a piece of software. A hash value, which is essentially a fingerprint of a file, is a numeric computation of a file's content which is used to identify the file. Two files with the same hash values are exact copies of each other.

⁵Rule references are to the Tax Court Rules of Practice and Procedure.

⁶Rule 72(a) provides:

(continued...)

In this context, documents and ESI include “writings, drawings, graphs, charts, photographs, sound recordings, images, and other data compilations stored in any medium from which information can be obtained, either directly or translated, if necessary, by the responding party into a reasonably usable form”.⁷ Rule 72(a)(1).

⁶(...continued)

RULE 72. PRODUCTION OF DOCUMENTS,
ELECTRONICALLY STORED INFORMATION, AND THINGS

(a) Scope: Any party may, without leave of Court, serve on any other party a request to:

(1) Produce and permit the party making the request, or someone acting on such party’s behalf, to inspect and copy, test, or sample any designated documents or electronically stored information (including writings, drawings, graphs, charts, photographs, sound recordings, images, and other data compilations stored in any medium from which information can be obtained, either directly or translated, if necessary, by the responding party into a reasonably usable form), or to inspect and copy, test, or sample any tangible thing, to the extent that any of the foregoing items are in the possession, custody, or control of the party on whom the request is served; * * *

⁷Literature on electronic data storage has characterized electronically stored data as falling within five categories. See Zubulake v. UBS Warburg LLC, 217 F.R.D. 309, 318 (S.D.N.Y. 2003). These categories are active, online data (e.g., hard drives); near-line data (e.g., optical disks); offline storage/archives (i.e., removable optical disk or magnetic tape media); backup tapes (i.e., a device that reads data from and writes it onto a tape); and fragmented, erased, or damaged data (fragmented data consists of files that are broken up and placed randomly throughout the disk). See id. at 318-319. The first three categories are generally considered accessible, while the remaining categories are generally considered inaccessible. See id. at 319-320.

And a party is generally required to produce documents or electronically stored information in the form in which they are maintained. Rule 72(b)(3). A party, however, is not required to provide discovery of ESI from sources that the party establishes are not reasonably accessible because of undue burden or cost unless the Court concludes that the requesting party has shown good cause for the discovery.⁸ See Rule 70(c)(2). These Rules are all similar to corresponding provisions found in the Federal Rules of Civil Procedure. See Fed. R. Civ. P. 34(a)(1)(A), (b)(2)(E), and 26(b)(2)(B).

II. Respondent's Request

Respondent requests access to petitioners' ESI. Petitioners resist this request, primarily because of cost and of concern that privileged or confidential information will be improperly disclosed. Respondent essentially responds that he can alleviate both concerns if petitioners give him all of the requested information, with a condition that he will allow them to later claim that some or all of that information should not be disclosed further because it is privileged. Petitioners remain mindful of their need to protect their privileged or confidential information, as well as the projected cost of protecting that information, and ask

⁸Petitioners do not claim that, if they use predictive coding, the requested ESI is not reasonably accessible because of undue burden or cost.

the Court to allow them to use predictive coding in responding to respondent's request.

In this respect, we note that this request is somewhat unusual. Our Rules are clear that "the Court expects the parties to attempt to attain the objectives of discovery through informal consultation or communication" before resorting to formal discovery procedures. Rule 70(a)(1). And although it is a proper role of the Court to supervise the discovery process and intervene when it is abused by the parties, the Court is not normally in the business of dictating to parties the process that they should use when responding to discovery. If our focus were on paper discovery, we would not (for example) be dictating to a party the manner in which it should review documents for responsiveness or privilege, such as whether that review should be done by a paralegal, a junior attorney, or a senior attorney. Yet that is, in essence, what the parties are asking the Court to consider--whether document review should be done by humans or with the assistance of computers. Respondent fears an incomplete response to his discovery. If respondent believes that the ultimate discovery response is incomplete and can support that belief, he can file another motion to compel at that time. Nonetheless, because we have not previously addressed the issue of computer-assisted review tools, we will address it here.

III. Expert Witnesses

Each party called a witness to testify at the evidentiary hearing as an expert. Petitioners' witness was James R. Scarazzo. Respondent's witness was Michael L. Wudke. The Court recognized the witnesses as experts on the subject matter at hand.

We may accept or reject the findings and conclusions of the experts, according to our own judgment. See Chapman Glen, Ltd. v. Commissioner, 140 T.C. 294, 329 (2013). We also may be selective in deciding what parts (if any) of their opinions to accept. See id.

IV. Analysis

The Court applies the standard of relevancy liberally when it comes to matters of discovery, see, e.g., Zaentz v. Commissioner, 73 T.C. 469, 471 (1979), and a party challenging the requested production of a document (including ESI) has the burden of establishing that the document is not discoverable, see Rutter v. Commissioner, 81 T.C. 937, 948 (1983); Branerton Corp. v. Commissioner, 64 T.C. 191, 192-193 (1975).

We believe that respondent's request for the ESI is within the bounds of our Rules, and petitioners do not appear to contest this point. At the same time, however, we are faced with the competing interests of the parties. On one hand,

we do not consider it appropriate to order petitioners to give all of their ESI to respondent, subject to a right to later claim that some or all of the information that he has reviewed is privileged or confidential information and thus outside the bounds of discovery. Although the use of a clawback agreement may be an option to which the parties might consent, petitioners reasonably resist entering into any such agreement as part of a plan under which they would voluntarily allow respondent to see all of the privileged or confidential information on the requested tapes. On the other hand, given the time and expense involved with petitioners' review of all the ESI to identify any privileged or confidential information, we likewise do not consider it appropriate to order petitioners to go to that extreme either.

We find a potential happy medium in petitioners' proposed use of predictive coding. Predictive coding is an expedited and efficient form of computer-assisted review that allows parties in litigation to avoid the time and costs associated with the traditional, manual review of large volumes of documents. Through the coding of a relatively small sample of documents, computers can predict the relevance of documents to a discovery request and then identify which documents are and are not responsive. The parties (typically through their counsel or experts) select a sample of documents from the universe of those documents to be searched

by using search criteria that may, for example, consist of keywords, dates, custodians, and document types, and the selected documents become the primary data used to cause the predictive coding software to recognize patterns of relevance in the universe of documents under review. The software distinguishes what is relevant, and each iteration produces a smaller relevant subset and a larger set of irrelevant documents that can be used to verify the integrity of the results. Through the use of predictive coding, a party responding to discovery is left with a smaller set of documents to review for privileged information, resulting in a savings both in time and in expense. The party responding to the discovery request also is able to give the other party a log detailing the records that were withheld and the reasons they were withheld.

Magistrate Judge Andrew Peck published a leading, oft-cited article on predictive coding which is helpful to our understanding of that method. See Andrew Peck, "Search, Forward: Will Manual Document Review and Keyboard Searches be Replaced by Computer-Assisted Coding?", L. Tech. News (Oct. 2011). The article generally discusses the mechanics of predictive coding and the shortcomings of manual review and of keyword searches. The article explains that predictive coding is a form of "computed-assisted coding", which in turn means "tools * * * that use sophisticated algorithms to enable the computer to determine

relevance, based on interaction with (i.e., training by) a human reviewer.” Id. at

29. The article explains that

Unlike manual review, where the review is done by the most junior staff, computer-assisted coding involves a senior partner (or team) who review and code a “seed set” of documents. The computer identifies properties of those documents that it uses to code other documents. As the senior reviewer continues to code more sample documents, the computer predicts the reviewer’s coding. (Or, the computer codes some documents and asks the senior reviewer for feedback.)

When the system’s predictions and the reviewer’s coding sufficiently coincide, the system has learned enough to make confident predictions for the remaining documents. Typically, the senior lawyer (or team) needs to review only a few thousand documents to train the computer.

Some systems produce a simple yes/no as to relevance, while others give a relevance score (say, on a 0 to 100 basis) that counsel can use to prioritize review. For example, a score above 50 may produce 97% of the relevant documents, but constitutes only 20% of the entire document set.

Counsel may decide, after sampling and quality control tests, that documents with a score of below 15 are so highly likely to be irrelevant that no further human review is necessary. Counsel can also decide the cost-benefit of manual review of the documents with scores of 15-50.

[Id.]

The substance of the article was eventually adopted in an opinion that states:

“This judicial opinion now recognizes that computer-assisted review is an

acceptable way to search for relevant ESI in appropriate cases.” Moore v. Publicis Groupe, 287 F.R.D. 182, 183 (S.D.N.Y. 2012), adopted sub nom. Moore v. Publicis Groupe SA, No. 11 Civ. 1279 (ALC)(AJP), 2012 WL 1446534 (S.D.N.Y. Apr. 26, 2012).

Respondent asserts that predictive coding should not be used in these cases because it is an “unproven technology”. We disagree. Although predictive coding is a relatively new technique, and a technique that has yet to be sanctioned (let alone mentioned) by this Court in a published Opinion, the understanding of e-discovery⁹ and electronic media has advanced significantly in the last few years, thus making predictive coding more acceptable in the technology industry than it may have previously been. In fact, we understand that the technology industry now considers predictive coding to be widely accepted for limiting e-discovery to relevant documents and effecting discovery of ESI without an undue burden.¹⁰

See Progressive Cas. Ins. Co. v. Delaney, No. 2:11-cv-00678-LRH-PAL, 2014 WL

⁹We use the term “e-discovery” to refer to “electronic discovery”, which in turn means the obtaining of ESI in the discovery phase of litigation.

¹⁰Predictive coding is so commonplace in the home and at work in that most (if not all) individuals with an email program use predictive coding to filter out spam email. See Moore v. Publicis Groupe, 287 F.R.D. 182, n.2 (S.D.N.Y. 2012), adopted sub nom. Moore v. Publicis Groupe SA, No. 11 Civ. 1279 (ALC)(AJP), 2012 WL 1446534 (S.D.N.Y. Apr. 26, 2012).

3563467, at *8 (D. Nev. July 18, 2014) (stating with citations of articles that predictive coding has proved to be an accurate way to comply with a discovery request for ESI and that studies show it is more accurate than human review or keyword searches); F.D.I.C. v. Bowden, No. CV413-245, 2014 WL 2548137, at *13 (S.D. Ga. June 6, 2014) (directing that the parties consider the use of predictive coding). See generally Nicholas Barry, “Man Versus Machine Review: The Showdown between Hordes of Discovery Lawyers and a Computer-Utilizing Predictive-Coding Technology”, 15 Vand. J. Ent. & Tech. L. 343 (2013); Lisa C. Wood, “Predictive Coding Has Arrived”, 28 ABA Antitrust J. 93 (2013). The use of predictive coding also is not unprecedented in Federal litigation. See, e.g., Hinterberger v. Catholic Health Sys., Inc., No. 08-CV-3805(F), 2013 WL 2250603 (W.D.N.Y. May 21, 2013); In Re Actos, No. 6:11-md-2299, 2012 WL 7861249 (W.D. La. July 27, 2012); Moore, 287 F.R.D. 182. Where, as here, petitioners reasonably request to use predictive coding to conserve time and expense, and represent to the Court that they will retain electronic discovery experts to meet with respondent’s counsel or his experts to conduct a search acceptable to respondent, we see no reason petitioners should not be allowed to use predictive coding to respond to respondent’s discovery request. Cf. Progressive Cas. Ins. Co., 2014 WL 3563467, at *10-*12 (declining to allow the use of predictive

coding where the record lacked the necessary transparency and cooperation among counsel in the review and production of ESI responsive to the discovery request).

Mr. Scarazzo's expert testimony supports our opinion.¹¹ He testified that discovery of ESI essentially involves a two-step process. First, the universe of data is narrowed to data that is potentially responsive to a discovery request. Second, the potentially responsive data is narrowed down to what is in fact responsive. He also testified that he was familiar with both predictive coding and keyword searching, two of the techniques commonly employed in the first step of the two-step discovery process, and he compared those techniques by stating:

[K]ey word searching is, as the name implies, is a list of terms or terminologies that are used that are run against documents in a method of determining or identifying those documents to be reviewed. What predictive coding does is it takes the type of documents, the layout, maybe the whispets of the documents, the format of the documents, and it uses a computer model to predict which documents out of the whole set might contain relevant information to be reviewed.

So one of the things that it does is, by using technology, it eliminates or minimizes some of the human error that might be associated with it. Sometimes there's inefficiencies with key word searching in that it may include or exclude documents, whereas training the model to go back and predict this, we can look at it and use statistics and other sampling information to pull back the

¹¹Mr. Wudke did not persuasively say anything to erode or otherwise undercut Mr. Scarazzo's testimony.

information and feel more confident that the information that's being reviewed is the universe of potentially responsive data.

He concluded that the trend was in favor of predictive coding because it eliminates human error and expedites review.

In addition, Mr. Scarazzo opined credibly and without contradiction that petitioners' approach to responding to respondent's discovery request is the most reasonable way for petitioners to comply with that request. Petitioners asked Mr. Scarazzo to analyze and to compare the parties' dueling approaches in the setting of the data to be restored from Dynamo's backup tapes and to opine on which of the approaches is the most reasonable way for petitioners to comply with respondent's request. Mr. Scarazzo assumed as to petitioners' approach that the restored data would be searched using specific criteria, that the resulting information would be reviewed for privilege, and that petitioners would produce the nonprivileged information to respondent. He assumed as to respondent's approach that the restored data would be searched for privileged information without using specific search criteria, that the resulting privileged information would be removed, and that petitioners would then produce the remaining data to respondent. As to both approaches, he examined certain details of Dynamo's backup tapes, interviewed the person most knowledgeable on Dynamo's backup

process and the contents of its backup tapes (Dynamo's director of information technology), and performed certain cost calculations.

Mr. Scarazzo concluded that petitioners' approach would reduce the universe of information on the tapes using criteria set by the parties to minimize review time and expense and ultimately result in a focused set of information germane to the matter. He estimated that 200,000 to 400,000 documents would be subject to review under petitioners' approach at a cost of \$80,000 to \$85,000, while 3.5 million to 7 million documents would be subject to review under respondent's approach at a cost of \$500,000 to \$550,000.

Our Rules, including our discovery Rules, are to "be construed to secure the just, speedy, and inexpensive determination of every case." Rule 1(d). Petitioners may use predictive coding in responding to respondent's discovery request. If, after reviewing the results, respondent believes that the response to the discovery request is incomplete, he may file a motion to compel at that time. See Rule 104(b), (d).

Accordingly,

An appropriate order will be
issued.