Pursuant to Tax Court Rule 50(f), orders shall not be treated as precedent, except as otherwise provided.

UNITED STATES TAX COURT
WASHINGTON, DC 20217

DRC

Dynamo Holdings Limited Partnership, Dynamo, GP, Inc., Tax Matters Partner, et al.,	)	
Petitioners,	)	
v.	)	Docket No. 2685-11, 8393-12.
COMMISSIONER OF INTERNAL REVENUE,	)	
Respondent	)	

O R D E R

These consolidated cases are calendared for trial at the special session of the Court beginning January 23, 2017, in Miami, Florida.

On July 25, 2013, the Commissioner filed a motion to compel the production of documents. The Commissioner requested that petitioners, Dynamo Holdings Limited Partnership and Beekman Vista, Inc., produce electronically stored information relating to adjustments and transfers between petitioners. Petitioners filed an objection arguing that if they had to produce the electronically stored information, they should be permitted to use predictive coding, a computer-based discovery tool, to respond.

On September 17, 2014, the Court issued Dynamo Holdings Limited P'ship v. Commissioner, 143 T.C. 183 (2014). In Dynamo, the Court stated its views on predictive coding. Id. at 190 ("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."). The Court granted the Commissioner's motion in that we compelled petitioners to produce the backup tapes from month end of August 2010 and month end of January 2008. Id. at 185. However, the Court allowed petitioners to respond using predictive coding. Id. at 194.

The quality of that response is now before us. Using a process described in more detail below, petitioners responded to the discovery requests by using

SERVED Jul 13 2016

predictive coding. The Commissioner, believing the response to be incomplete, served petitioners with a new discovery request asking for all documents containing any of a series of search terms. (Those same search terms had been used in a Boolean search during the predictive coding process to identify how many documents in the electronic records had each term.) Petitioners objected to this new discovery request as duplicative of the previous discovery responses made through the use of predictive coding. On June 17, 2016, the Commissioner filed a motion under Tax Court Rules 72(b)(2)[1] to compel the production of documents responsive to the Boolean search that were not produced through the use of predictive coding. The petitioners object.

Discussion

When responding to a document request, technology has rendered the traditional approach to document review impracticable. The traditional method is labor intensive, with people reviewing documents to discern what is (or is not) responsive, with the responsive documents then reviewed for privilege, and with the responsive and non-privileged documents being produced. When reviewing documents in the dozens, hundreds, or low thousands, this worked fine. But with the advent of electronic recordkeeping, documents no longer number in the mere thousands, and various electronic search methods have developed.

When electronic records are involved, perhaps the most common technique that is employed is to begin with keyword searches or Boolean searches to a defined universe of documents. Then, the responding party typically reviews the results of those searches to identify what, in fact, is responsive to the request. Implicit in this approach is the fact that some of the documents that are responsive to the word or Boolean search are responsive, while others are not.

An emerging approach, and the approach authorized in this case in our Opinion at 143 T.C. 183, is to use predictive coding to identify those documents that are responsive. A few key points of that Opinion are worth highlighting.

First, the Court authorized the responding party (petitioners) to use predictive coding, but the Court did not, in either its Opinion or its subsequent Order of September 17, 2014, mandate how the parties proceed from that point.

---

[1] Unless otherwise indicated, all Rule references are to the Tax Court Rules of Practice and Procedure.

The parties are to be commended for working together to develop a predictive coding protocol from which they worked.

Second, the Court held open the issue of whether the resulting document production would be sufficient, expressly stating "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." Id. at 189, 194.

To state the obvious, (1) it is the obligation of the responding party to respond to the discovery, and (2) if the requesting party can articulate a meaningful shortcoming in that response, then the requesting party can seek relief. We turn now to those two points.

The Predictive Coding Response

First, we address what petitioner did to respond to the discovery. We draw this factual background from the Commissioner's motion to compel and petitioners' responses, the attached exhibits, the parties' 16 joint status reports, the Court's orders, and the Court's June 30, 2016 conference call with the parties.

As indicated by the parties 16 joint status reports, the parties generally agreed to and followed a framework for producing the electronically stored information using predictive coding: (1) restoring and processing the backup tapes, (2) selecting and reviewing seed sets, (3) establishing and applying the predictive coding algorithm; and (4) reviewing and returning the production set.

First, the parties agreed on how to restore and process the two backup tapes from month end of August 2010 and month end of January 2008. The parties agreed that petitioners would restore the two tapes. While petitioners were restoring the first backup tape, the Commissioner requested that petitioners conduct a Boolean search and provided petitioners with a list of search terms for petitioners to run against the processed data. The list included 76 search terms, which were categorized into persons, property transfers, amounts, adjustments, and documents of interest.

While petitioners were processing the second tape, petitioners notified the Commissioner that an Exchange database file, the main file containing petitioners' e-mails, was not backed up on the second tape. Petitioners located the Exchange

database. The Commissioner requested that petitioners run the Boolean search using the same search terms against the processed data.

Petitioners conducted a Boolean search for the 76 search terms on the first tape and the Exchange database tape. Petitioners searched 406,939 documents and provided the Commissioner with a table of the results. The table included "individual term hits," "documents with term hits", and "individual documents only containing the single term." For example, the term "196,967,422 OR 196967422" had 52 hits in a total of 9 documents, and 3 of those 9 documents contained only that term (meaning that 6 documents contained that term plus at least one other search term). Likewise, the term "21,249,810 OR 21249810" had 9 hits in one document and there were no documents containing only that term.

Second, the parties agreed on how to select and review the seed sets. Initially, the Commissioner requested the seed sets be selected from the documents containing hits from the Boolean search. However, the parties agreed that petitioners would randomly select two sets of 1,000 documents from the first tape and the Exchange database tape. After receiving the seed sets, the Commissioner identified which documents were relevant or not relevant, and this coding trained the predictive coding model to identify responsive documents. That model was then used to test how well it performed on the second seed set of 1,000 that the Commissioner coded. The Commissioner wanted to train the predictive coding model so that it returned 95 percent of the relevant documents he identified on the second set of 1,000.

After the model was run against the second 1,000 documents, petitioners' technical professionals reported that the model was not performing well. The parties agreed that the Commissioner would code additional sets of documents. The parties agreed that petitioners would provide to the Commissioner 10 additional sets of approximately 100 documents that were richer in relevant material to make the training process more productive than the random sample. After the Commissioner completed coding the additional sets, the parties also determined that the Commissioner would review 2 additional sets of 100 documents that the model currently coded as having a high relevancy score. After the Commissioner completed his review of the additional sets of 100 documents, petitioners' technical professionals suggested that the parties could consider a final validation sample of 1,000 documents to test the performance model, but they explained that the additional review would unlikely improve the model. The Commissioner declined to code a final validation set.

Third, the parties agreed that the Commissioner would establish a recall rate for the predictive coding algorithm and it would be applied to the backup tapes. Petitioners provided the Commissioner with a table summarizing the anticipated return results. The table showed that the model could generate a range of recalls between 65 percent to 95 percent. The higher the selected recall rate, the greater the amount of relevant and nonrelevant documents produced. The Commissioner selected a 95 percent recall rate. After the Commissioner selected the recall rate, the predictive coding algorithm was ready to be applied to the backup tapes.

The parties could not agree on the final step of reviewing and returning the production set. On December 15, 2015, the Court entered an agreed order directing the parties on the remaining steps to complete petitioners' delivery of the production set, and the Commissioner's review and return of that information.

In accordance with the Court's order, petitioners ran the algorithm to identify documents at a 95 percent recall rate against the initial set of 2 backup tapes, approximately 406,000 documents. Petitioners then ran a second algorithm on the initial set to identify privileged materials. On January 4, 2016 and March 3, 2016, petitioners delivered a production set of approximately 180,000 total documents on a portable device for the Commissioner to review ("the production set"). Petitioners included a relevancy score for each document.

After the Commissioner reviewed the documents, he retained 5,796 documents ("the retained documents") and returned to petitioners the remaining documents. The Commissioner provided petitioners with a list of the retained documents.

Alleged Shortcomings of the Response

On June 17, 2016, the Commissioner filed a motion to compel production of the documents identified in the Boolean search that were not produced in the production set. The Commissioner speculates that these documents are "highly likely to be relevant." In the Commissioner's motion, he includes a table with 23 terms that were used in the Boolean search. The Commissioner asserts that when petitioners conducted the Boolean search of these terms, there were 1,645 documents containing those terms; but when petitioners delivered the production set, 1,353 of those documents were excluded.

On June 27, 2016, petitioners filed an objection to the Commissioner's motion to compel. Petitioners contend that the predictive coding algorithm worked correctly, and the Commissioner's calculations are wrong. Petitioners observe that only 1,360 documents (not 1,645) contain those terms because some documents contain more than one search term. Further, petitioners allege that 440 of the 1,360 documents were produced (with 13 clawed back as privileged), bringing the universe of documents at issue down to 920. Petitioners contend that 765 documents were excluded as not relevant by the predictive coding algorithm. Petitioners speculate that these documents were excluded because they are outside the relevant time frame or otherwise are not relevant. In any event, petitioners argue that the documents were selected by the predictive coding algorithm based on selection criteria set by the Commissioner.

The Court held a conference call with the parties to gain a better understanding of these facts and offered the parties an opportunity to supplement their papers. On July 5, 2016, petitioners filed a supplement to their objection. In their supplement, petitioners explain that in sampling the 765 documents that were not produced, they found that many of the documents predate or postdate the relevant time period. Petitioners also explain that the Commissioner is incorrect for at least one of the terms, "21,249,810 OR 21249810" (the first term on the Commissioner's list).[2] According to the Boolean search, there was one document containing this term. According to the Commissioner, this document had not been produced. Petitioners contend that this document was not only produced as part of the production set, but also that it was among the documents that the Commissioner selected and retained. Petitioners identified the document by Bates number.

Recall Versus Precision

Before moving on, it is helpful to define two concepts relevant to searching and retrieving documents: recall and precision.

> A search method's precision is defined as the percentage of documents retrieved by the methods that are relevant. The higher a search's precision, the fewer "false positives" there are. A search method's recall is defined as the percentage of all relevant documents in the search universe that are retrieved by that search method. The higher the recall, the fewer "false

[2] This search term had been used by the Court in its June 30, 2016 conference call with the parties as an example for gaining a better understanding of the searches that had been conducted.

> negatives" (i.e., relevant but unretrieved documents) there are. Often, there is a trade-off between precision and recall—a broad search that misses few relevant documents will usually capture a lot of irrelevant documents, while a narrower search that minimizes "false positives" will be more likely to miss some relevant documents.

L-3 Commc'ns Corp. v. Spartons Corp., 313 F.R.D. 661, 666-667 (M.D.Fla. 2015) (internal cites omitted) (citing The Sedona Conference Best Practices Commentary on the Use of Search & Information Retrieval Methods in E-Discovery, 15 Sedona Conf. J. 217, 237 (2014)); see also Maura R. Grossman & Gordon V. Cormack, Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review, Rich. J.L. & Tech., Spring 2011, at 8-9.

Those numbers are often in tension with each other: as the predictive coding model is instructed to return a higher percentage of responsive documents, it is likely also to include more nonresponsive documents. Thus, when setting the recall rate at 95%, the Commissioner likewise chose a model that would return more nonresponsive documents (in this case, a precision rate of 3%).

Respondent (in effect) argues that the absence of some of the documents found in the Boolean search from the body of retained documents shows that the predictive coding response was flawed, or using the terms just defined, that its level of recall was too low. We will assume that it was flawed, but the question remains whether any relief should be afforded.

Respondent's motion is predicated on two myths.

The first is the myth of human review. As noted in The Sedona Conference Best Practices Commentary on the Use of Search & Information Retrieval Methods in E-Discovery: "It is not possible to discuss this issue without noting that there appears to be a myth that manual review by humans of large amounts of information is as accurate and complete as possible – perhaps even perfect – and constitutes the gold standard by which all searches should be measured." 15 Sedona Conf. J. 214, 230 (2014). This myth of human review is exactly that: a myth.

Research shows that human review is far from perfect. Several studies are summarized in Nicholas M. Pace & Laura Zakaras, RAND Corp., Where the

Money Goes: Understanding Litigant Expenditures for Producing Electronic Discovery (2012) at 55. To summarize even further, if two sets of human reviewers review the same set of documents to identify what is responsive, research shows that those reviewers will disagree with each on more than half of the responsiveness claims. As the RAND report concludes:

> Taken together, this body of research shows that groups of human reviewers exhibit significant inconsistency when examining the same set of documents for responsiveness under conditions similar to those in large-scale reviews. Is the high level of disagreement among reviewers with similar backgrounds and training reported in all of these studies simply a function of the fact that determinations of responsiveness or relevance are so subjective that reasonable and informed people can be expected to disagree on a routine basis? Evidence suggests that this is not the case. Human error in applying the criteria for inclusion, not a lack of clarity in the document's meaning or ambiguity in how the scope of the production demand should be interpreted, appears to be the primary culprit. In other words, people make mistakes, and, according to the evidence, they make them regularly when it comes to judging relevance and responsiveness.

Id. at 58. (Indeed, even keyword searches are flawed. One study summarized in Moore v. Publicis Groupe & MSL Grp., 287 F.R.D. 182, 191 (S.D.N.Y. 2012), found that the average recall rate based on a keyword review was only 20%.)

The second myth is the myth of a perfect response. The Commissioner is seeking a perfect response to his discovery request, but our Rules do not require a perfect response. Instead, the Tax Court Rules require that the responding party make a "reasonable inquiry" before submitting the response. Specifically, Rule 70(f) requires the attorney to certify, to the best of their knowledge formed after a "reasonable inquiry," that the response is consistent with our Rules, not made for an improper purpose, and not unreasonable or unduly burdensome given the needs of the case. Rule 104(d) provides that "an evasive or incomplete * * * response is to be treated as a failure to * * * respond." But when the responding party is signing the response to a discovery demand, he is not certifying that he turned over everything, he is certifying that he made a reasonable inquiry and to the best of his knowledge, his response is complete.

Likewise, "the Federal Rules of Civil Procedure do not require perfection." Moore, 287 F.R.D. at 191. Like the Tax Court Rules, the Federal Rule of Civil

Procedure 26(g) only requires a party to make a "reasonable inquiry" when making discovery responses.

The fact that a responding party uses predictive coding to respond to a request for production does not change the standard for measuring the completeness of the response. Here, the words of Judge Peck, a leader in the area of e-discovery, are worth noting:

> One point must be stressed - it is inappropriate to hold TAR [technology assisted review] to a higher standard than keywords or manual review. Doing so discourages parties from using TAR for fear of spending more in motion practice than the savings from using from using TAR for review.

Rio Tinto PLC v. Vale S.A., 306 F.R.D. 125, 129 (S.D.N.Y. 2015).

Conclusion

There is no question that petitioners satisfied our Rules when they responded using predictive coding. Petitioners provided the Commissioner with seed sets of documents from the backup tapes, and the Commissioner determined which documents were relevant. That selection was used to develop the predictive coding algorithm. After the predictive coding algorithm was applied to the backup tapes, petitioners provided the Commissioner with the production set. Thus, it is clear that petitioners satisfied our Rules with their response. Petitioners made a reasonable inquiry in responding to the Commissioner's discovery demands when they used predictive coding to produce any documents that the algorithm determined was responsive, and petitioners' response was complete when they produced those documents. Accordingly, it is

ORDERED that respondent's Motion to Compel Production of Documents Containing Certain Terms filed June 17, 2016 is denied.

(Signed) Ronald L. Buch
Judge

Dated: Washington, D.C.
July 13, 2016