WASHINGTON JOURNAL OF LAW, TECHNOLOGY & ARTS
VOLUME 8, ISSUE 4 WINTER 2013

UNCHAINING E-DISCOVERY IN THE PATENT COURTS

Daniel B. Garrie, Esq. and Yoav M. Griver, Esq.[*]

Cite as: 8 WASH. J.L. TECH. & ARTS 487 (2013)
http://digital.law.washington.edu/dspace-law/handle/1773.1/1228

ABSTRACT

This Article analyzes the Federal Circuit's Model Order Regarding E-Discovery in Patent Cases (the "Model Order"). The Article briefly describes the purpose behind the Model Order, describes its key provisions, analyzes the Model Order to identify some areas of continuing concern, and defines predictive coding to examine the impact, or lack thereof, on the Model Order. The Author concludes that, while it is beyond refute that the Model Order is an appropriate step toward controlling and managing e-discovery, the Model Order is only the first step. In this regard, several problems, as set forth below, can potentially arise when counsel or the courts use the Model Order. It is hoped that this Article will encourage judges, litigants, and other interested parties to continue trying to solve the continuously troubling aspects of e-discovery and e-discovery abuse.

---

[*] Daniel B. Garrie, Esq. is a partner at Law & Forensics, splitting his time between the East and West Coast offices, where he focuses on forensics, e-discovery, cyber security, and related investigations, including in the area of intellectual property disputes. Mr. Garrie can be reached at daniel@lawandforensics.com.

Yoav M. Griver, Esq. is a partner in the New York office of Zeichner Ellman & Krause LLP, where he concentrates his practice on complex litigation, technology, and intellectual property disputes. Mr. Griver can be reached at ygriver@zeklaw.com.

TABLE OF CONTENTS

INTRODUCTION

The Model Order Regarding E-Discovery in Patent Cases[1] ("Model Order") is the Federal Circuit's response to the exponential growth of e-discovery and related costs in cases before it.[2] As noted in the Introduction to the Model Order, patent cases tend to suffer from disproportionally high discovery expenses—one study determined that the costs of Intellectual Property cases run almost 62 percent more than other litigation.[3] Moreover, the exponential growth in electronic documents and communications has, intentionally or otherwise, led to what the Federal Circuit considers to be excessive e-discovery.[4] Broad and unfettered e-discovery—particularly email-related discovery—facilitated litigation where the time and cost of electronic production far outweighed the minimal benefits of marginal and cumulative disclosure, thus threatening to derail the judicial promise of just, speedy, and affordable dispute resolution:

> As technology and knowledge play an increasingly important role in our economy, litigation must not become an intolerably expensive way to resolve patent disputes. Specifically, litigation costs should not be permitted to unduly interfere with the availability of the court to those who seek to vindicate their patent rights. The enforcement of such rights is both an obligation of the legal system and important to innovation. Likewise, disproportionate expense should not be permitted to force those accused of infringement to acquiesce to nonmeritorious claims.[5]

The Model Order provides the courts and counsel with a

---

[1] Fed. Cir., E-Discovery Committee, Model Order Regarding E-Discovery in Patent Cases, available at http://www.cafc.uscourts.gov/images/stories/the-court/Ediscovery_Model_Order.pdf [hereinafter Model Order].

[2] Fed. Cir., E-Discovery Committee, Introduction to Model Order Regarding E-Discovery in Patent Cases, 2, available at http://www.cafc.uscourts.gov/images/stories/the-court/Ediscovery_Model_Order.pdf [hereinafter Introduction to Model Order].

[3] *See id.* at 1 (citing Emery G. Lee III & Thomas E. Willging, *Litigation Costs in Civil Cases: Multivariate Analysis* 8 (Fed. Judicial Ctr. 2010)).

[4] *Id.* at 2.

[5] Introduction to Model Order at 2.

framework to manage the e-discovery process, particularly for the responsible and targeted use of e-discovery in patent cases. It seeks to "promote economic and judicial efficiency by streamlining e-discovery, particularly email production, and requiring litigants to focus on the proper purpose of discovery—the gathering of material information . . . ."[6]

I. A REVIEW OF THE KEY PROVISIONS OF THE MODEL ORDER

The Model Order attempts to initiate mutual targeted e-discovery by placing presumptive limits on e-discovery. In this regard, the Model Order patterned itself after Federal Rule of Civil Procedure 30, which limited deposition practice by presumptively limiting each side to ten depositions of seven hours each.[7] Specifically, the Model Order requires the parties exchange the type of core documentation key to every patent litigation—i.e., documents concerning (i) the patent, (ii) the accused product, (iii) the prior art, and (iv) the relevant finances—before propounding email requests.[8] Even then, the Model Order presumptively limits the number of custodians and search terms for all email production requests, such that any email production requests remain focused on appropriate e-discovery issues.[9] These limits are presumptive only, and may be modified by the parties or the court for good cause shown.[10]

Where a party seeks more discovery than agreed upon by the parties, or allowed by the court, the requesting party bears the reasonable cost of that discovery.[11] By shifting costs, the Model Order seeks to ensure that a party carefully balances the cost and value of the additional discovery.[12]

The Model Order also seeks to lower the cost of e-discovery by

---

[6] *Id.*; *see also* Model Order ¶ 1 ("This Order . . . streamlines Electronically Stored Information ('ESI') production to promote a 'just, speedy, and inexpensive determination' of this action.").

[7] *See id.* at 3; FED. R. CIV. P. 30.

[8] Model Order ¶ 8.

[9] *Id.* ¶¶ 6, 7, 10, 11.

[10] *Id.* ¶ 2.

[11] *Id.* ¶¶ 10, 11.

[12] *Id.*; Introduction to Model Order at 3-4 .

addressing a large source of that cost—pre-production document review by attorneys or other human reviewers. To minimize this pre-production review, the Model Order expressly provides that the inadvertent production of work product or attorney-client privileged work during the e-discovery period may not be used in the pending case. Moreover, these protected works neither constitute a waiver in the pending case, nor in any other federal or state proceeding.[13]

II. THE MODEL ORDER: AREAS OF CONTINUING CONCERN

The Model Order is a good first step toward addressing the major problem with e-discovery: ever-increasing complexity and expense. However, the Model Order solutions raise several concerns, four of which are identified and discussed below.

A. The Model Order's Cost-Shifting Triggers Allow the Parties to "Game" the System and May Offer Disincentives to More Economical Alternatives in E-Discovery

The first potential area of concern regarding the Model Order arises from the Model Order's reliance on disproportionate costs that trigger cost-shifting.[14] It is possible for the producing party's counsel to manipulate the discovery process so as to increase costs and force the requesting party to bear those costs. Specifically, the costs of performing data collection or execution may sometimes be substantially less costly if done in-house than if a third-party vendor collected and performed the search. For example, a large technology firm might have a proprietary document-tracking platform that runs on legacy hardware and an in-house IT team managing this system. In such cases, it would be substantially more costly to retain a third-party vendor than to use the in-house IT department. Yet, that expense arguably could still be presented

---

[13] Model Order ¶¶ 12-14; Introduction to Model Order at 4.

[14] Model Order ¶ 3. The Model Order also provides that discovery tactics that delay or prolong the process will be considered by the court in determining which party should bear the costs of the discovery process. *Id.*

to the court and opposing counsel as a true cost in e-discovery, consequently deterring, narrowing, or shifting e-discovery costs. Indeed, the producing party can contend that using a third-party vendor is appropriate because doing so will avoid any concern that in-house IT staff will inevitably skew the production results in favor of the producing party. The end result is that a party can, or at least can try, to intentionally trigger cost-shifting as a litigation tactic.

Courts and litigants should be aware of this tactic, and raise the issue during the initial discovery conference mandated by Federal Rule of Civil Procedure 26. One solution involves the courts encouraging parties to utilize their own IT departments when possible to collect and produce documents, as long as "best practices" are followed by the in-house IT department in collecting and producing those documents.

B. The Model Order Default Standard That Metadata is not to be Produced Absent a Showing of Good Cause Ignores the Critical Value Metadata Provides When Issues Exist Around Authenticity or Authorship

The second area of concern with the Model Order is its default standard allowing no metadata (i.e., "data about data") absent a showing of good cause.[15] In patent-related disputes focusing on the patent creator's identity, filing date, or general priority, metadata is likely to be a critical element in the discovery process; metadata access can yield critical information regarding such key points as dates, times, authorship, and other related elements.[16] Although the Model Order does allow parties to request metadata upon a showing of good cause, it is an uphill effort for counsel to establish good cause because litigants may not have enough information to specifically determine what metadata they need in order to make the necessary showing.

One solution is for a court to maintain a lenient standard for

---

[15] Model Order ¶ 5.

[16] *See* DANIEL B. GARRIE & YOAV M. GRIVER, DISPUTE RESOLUTION AND E-DISCOVERY (2012).

good cause, and allow relevant facts to emerge early in the case to reserve litigation time and cost.

C. The Model Order Only Allows Email Production to Occur After the Parties Have Exchanged Initial Disclosures of Basic Documents and Information on the Critical Systems Storing the Email

The Model Order attempts to force the parties to hold off on email production until after initial disclosures regarding relevant financial information, prior art, and patents.[17] However, to encourage focused and reasonable e-mail production, the Model Order also should require the parties to define their respective technology systems involved with email. This information is critical to drafting reasonable and narrowly tailored email requests required by the Model Order.[18]

For example, a party might craft an e-mail request that is narrowly tailored and appears reasonable,[19] but that request may be unreasonable if the party seeks email that is five years old and remains stored on disk backup in Germany. In this example, the cost of production, given the medium and location, makes an apparently narrow and reasonable request unreasonable in practice, and may require an even more refined request.

The parties should be required to identify and disclose their respective technology systems involved with email as a preliminary matter, so that such issues may be identified before

---

[17] Model Order ¶ 8.

[18] *See* Model Order ¶ 6 ("To obtain email parties must propound specific email production requests."); ¶ 7 ("Email production requests shall only be propounded for specific issues, rather than general discovery of a product or business.").

[19] *See*, *e.g.*, McGrath v. United States, No. 11–318C, 2012 WL 726423 (Fed. Cl. March 6, 2012). In *McGrath*, the United States Court of Federal Claims considered a proposed discovery order that contained some, but not all, of the provisions from the Model Order. Among other things, the parties were eventually ordered to cooperate to identify the proper custodians, proper search terms, and proper timeframe before producing email, and "encouraged" to use narrowing search criteria (e.g., "and," "but not," "w/x") to limit email production.

each party issues email requests. Thus, one possible solution is a Model Order amendment requiring the parties to exchange information about their IT systems at the earliest litigation stage, enabling both sides to effectively organize their forthcoming search requests.

D. The Model Order Should Consider Requiring the Parties to Perform Email Sampling Before Limiting the Number of Search Terms and Custodians to Five People and Terms

The Model Order presumptively limits the number of custodians and search terms for all email production requests to five terms and custodians per producing party.[20] The intent is to control exorbitant production costs by minimizing what parties can request.[21] Although well-intentioned, this presumptive limit presents a challenging paradigm; it is impossible for parties to be 100 percent accurate on terms and custodians, especially when they do not control the data. Consequently, prior to selecting terms or custodians, the court or parties should filter the available field with common sense:

1. Both parties should group search terms into high-, medium-, and low-value groups.
2. The parties should then take each group of search terms and identify applicable timeframes and custodians.

For example:

> High Group
> Dates: 02/2010 to 05/2011; 03/2005 to 04/2006
> Custodians: D. Smith; M. Jane
> Terms: Apple, Democrat, Republican, Libertarian

3. The opposing party should then sample each of the custodians using the search terms and dates for the group.
4. The proposing party may then re-order the terms and custodians.

---

[20] Model Order ¶¶ 10, 11.
[21] Introduction to Model Order at 2; Model Order ¶¶ 6, 7.

Of course, the court should mandate applying the Model Order's strict number requirements if the parties fail to mutually agree on a protocol, or if the terms the parties propose are inappropriate or indiscriminate in nature. Under such circumstances, ¶ 11 of the Model Order provides for cost-shifting to the requesting party.

III. IMPACT OF PREDICTIVE CODING ON THE FEDERAL CIRCUIT MODEL ORDER

Currently, there is a debate over the effectiveness and reliability keyword searching,[22] as many contend that there are superior ways to return the most responsive documents in a litigation matter.[23] Bolstering this contention is predictive coding, or technology-assisted review,[24] which increasingly strengthens

---

[22] *See* Victor Stanley, Inc. v. Creative Pipe, Inc., 250 F.R.D. 251, 259-60 (D. Md. 2008). *See generally* Da Silva Moore v. Publicis Groupe, 868 F. Supp.2d 137 (S.D.N.Y. 2012) (order approving use of computer-assisted review). While there is a great deal of debate, condemning keyword search technology is unfair in this context, as it is akin to saying tape players are inferior to MP3 players, and predictive coding at various stages heavily relies upon the selection of the proper keywords.

[23] Technologies like de-duplication and keyword searching are used to limit the volume of documents reviewed in the discovery process and predictive coding software is just another piece of software; it just so happens that predictive coding limits the volume of documents by a greater multiple than other solutions available in the marketplace. Maura R. Grossman & Gordon V. Cormack, *Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review*, XVII RICH. J.L. & TECH. 11 (2011). *See generally*, Bruce Hedin et al., Conference Report, *Overview of the TREC 2009 Legal Track*, SP 500-278 NIST Special Publ'n: 18th Text REtrieval Conf. Proc. (2009), http://trec.nist.gov/pubs/trec18/papers/LEGAL09.OVERVIEW.pdf.; Douglas W. Oard et al., Conference Report, *Overview of the TREC 2008 Legal Track*, SP 500-277 NIST Special Publ'n: 17th Text REtrieval Conf. Proc. (2008), http://trec.nist.gov/pubs/trec17/papers/LEGAL.OVERVIEW08.pdf.

[24] There is no agreement today as to what is the proper nomenclature for predictive coding technology. *See* Sharon D. Nelson and John W. Simek, *Predictive Coding: A Rose by any Other Name*, SENSEI ENTERPRISES, INC., http://www.senseient.com/storage/articles/Predictive_Coding.pdf (last visited Jan 24, 2013).

prospective alternatives to keyword search efficacy. While still somewhat contentious, courts have recognized that predictive coding software improves document review quality and efficiency and is a defensible means of facilitating discovery.[25] With these new tools in mind, the Model Order already requires amendment, as predictive coding mechanisms do not currently align with the Model Order's provisions.

A. What is Predictive Coding? How Does Predictive Coding Work?

Predictive coding is software trained by a user to predict which documents in a document set will be responsive and which will be non-responsive. In recent history, predictive coding has distinguished itself as an alternative to keyword searches for finding relevant documents within the ever-increasing document sets produced during discovery.[26] Predictive coding aims to reduce the number of documents reviewed by ranking available documents according to a calculated level of responsiveness.[27] Instead of looking at every email written by a custodian over a three-year time period, predictive coding uses numerous factors including subject matter, punctuation style, writing style, and

---

[25] Moore v. Publicis Groupe SA, 2012 WL 1446534, *3 (S.D.N.Y. Apr. 26, 2012) ("Manual review with keyword searches is costly, though appropriate in certain situations."). Computer-assisted review need not be used in all cases. Moore v. Publicis Groupe, 18 Wage & Hour Cas. 2d (BNA) 1479, 2012 WL 607412, *12 (S.D.N.Y. 2012), *adopted*, 2012 WL 1446534 (S.D.N.Y. 2012); *see also* A.N.S.W.E.R. Coal. v. Salazar, 258 F.R.D. 36, 38 (D.D.C. 2009) (ordering parties to confer on "a methodology for [keyword] searches [and] . . . a list of search directives that are likely to result in [relevant] documents"); Am. Family Mut. Ins. Co. v. Gustafson, No. 08-cv-02772-MSK-MJW, 2009 WL 641297, *3 (D. Colo. Mar. 10, 2009) (ordering parties to "meet, confer, and agree upon the search terms that will be used" to search imaged hard drive).

[26] *See generally* Jason R. Baron, *Law in the Age of Exabytes: Some Further Thoughts on 'Information Inflation' and Current Issues in E-Discovery Search*, XVII RICH. J.L. & TECH. 9 (2011); Daniel B. Garrie & Edwin A. Machuca, *E-Discovery Mediation & the Art of Keyword Search*, 13 CARDOZO J. CONFLICT RESOL. 467 (2012).

[27] *See generally Moore*, 2012 WL 607412 (describing function and use of predictive coding software).

keywords to determine the chain of documents most relevant to the particular issue or topic.[28] The underlying programmable algorithms will vary between software brands.[29]

Technically speaking, predictive coding is a computerized process that uses "sophisticated algorithms to enable the computer to determine relevance, based on interaction with (i.e., training by) a human reviewer".[30] In other words, the software is trained by a senior attorney or partner to look for documents similar to documents that the training attorney deems responsive. This is done by the attorney reviewing a relatively small "seed set" of documents, and coding the documents as either responsive or not responsive.[31] "The computer identifies properties of those documents that it [then] uses to code other documents" until the "system's predictions and the reviewer's coding sufficiently coincide," at which point "the system has learned enough to make confident predictions for the remaining documents."[32]

Typically, this procedure allows a set of hundreds of thousands of documents (or more) to be coded for responsiveness by the software, even though only a few thousand have actually undergone examination by a senior attorney or partner.[33] In a common implementation of the technique, documents coded "non-responsive" by the software—typically the bulk of any document collection—may never be examined again, except for quality-control sampling to ensure that the software's responsiveness interpretation matches that of the attorney running the software.[34] Those documents coded by the software as "responsive" are reviewed by attorneys for a final responsiveness determination, as

---

[28] *Id.*

[29] Alaap B. Shah, *Use of "Predictive Coding" to Limit Cost and Improve Efficiency in Healthcare E-discovery: The Light Is Green, But Proceed With Caution*, AHLA HEALTHCARE LIABILITY & LITIGATION, Jan. 13, 2012.

[30] *Moore*, 2012 WL 607412, *2.

[31] *Id.*

[32] *Id.*

[33] *See generally* Monique da Silva Moore v. Publicis Groupe, 868 F. Supp.2d 137 (S.D.N.Y. 2012); Daniel B. Garrie, *Effective Keyword Selection Requires a Mastery of Storage Technology and the Law*, 32 PACE L. REV. 400 (2012).

[34] *See generally Moore*, 2012 WL 607412.

well as for potentially privileged information.

When using predictive coding, counsel should understand how the software addresses recall, precision, and accuracy, as early computer-assisted review projects likely will require additional education for the reviewing court.[35] Counsel should understand the following concepts when assessing predictive coding technology:

- *Recall*[36] = Number of Documents Predicted to be Responsive / Total Number of Actually Responsive Documents.
- *Precision*[37] = Number of Actually Responsive Documents / Number of Documents Predicted to be Responsive.
- *Accuracy*[38] = (True Relevant Documents Retrieved + True Non-Relevant Documents Not Retrieved) / Total Documents.

There is a give-and-take relationship between Recall and Precision.[39] If recall is high (very broad search criteria), precision will be low (over-inclusive results). Conversely, if precision is high (stringent search criteria), recall will be low (under-inclusive results).[40] Richness is another concept to be aware of when looking at predictive coding statistics.[41] Richness is simply the percentage of relevant documents in the total population. It is a measure

---

[35] Da Silva Moore v. PUBLICIS GROUPE & MSL GROUP, No. 11 Civ. 1279 (ALC)(AJP) (S.D.N.Y. Feb. 24, 2012).

[36] Daniel B. Garrie, *Predictive Coding Endorsed by the Southern District of New York – People Plus Technology Equals Smart Search*, DAILY JOURNAL, May 1 2012.

[37] *Id.*

[38] *Id.*

[39] *Id.*

[40] *Id.* The F-score is intended to account for this trade-off. It is a statistical measure used in text retrieval applications specifically to measure the compromise between precision and recall. $F = 2(P*R)(P+R)$.

For example, if our recall is 80 percent, but our precision is only 10 percent, the F-score is 17 percent. This is sometimes known as the Balanced F-score or F_1 measure because it gives equal weight to both recall and precision. The F-score can be weighted if more emphasis needs to be placed on recall (the F_2 measure) or on precision (the $F_{0.5}$ measure), according to user requirements.

[41] *Id.*

related to the population, not to the computer's overall document review performance. Thus, in a collection containing few relevant documents across a large document population, such as evidence of well-hidden corporate criminal activity, the richness will be low because only a few relevant documents are returned. Conversely, in, for example, a medical malpractice suit, almost all of the documents produced by the medical facility regarding a particular patient will be responsive, which will result in a higher richness score.

B. Impact of Predictive Coding on the Bar

While predictive coding software may lead to disappointment for those first-year attorneys enamored with the excitement and challenges of document review, it will be the savior to many others.[42] The effectiveness of predictive coding software hinges on the initial software training performed by either a senior associate or a partner who is intimately aware of the underlying facts and litigation strategies accompanying a particular case.[43] While the notion of learning how to train associates on such a system may be overwhelming, training concerns are misplaced, as some predictive coding software is operationally intuitive. In the end, predictive coding lowers costs, saves times, and, if done properly, generates revenue while increasing the quality of professional life for partners and associates.[44]

---

[42] *See, e.g.*, Ralph Losey, *Bottom Line Driven Proportional Review*, E-DISCOVERY TEAM (Nov. 1, 2012), http://e-discoveryteam.com/2012/01/15/bottom-line-driven-proportional-review/ ("[Y]ou *cannot* just dispense with final manual review. . . . [W]e are not going to turn that over to the Borg anytime soon. I've asked around and no law firms do that now. No experts advocate that approach either, even the most extreme advocates for automation (of which I'm one). . . . You use predictive coding to speed up the final manual review to be sure, but only a fool (or con artist trying to get at a producing parties [sic] secrets) trusts coding software today without human verification.").

[43] *See generally* Alaap B. Shah, *Use of "Predictive Coding" to Limit Cost and Improve Efficiency in Healthcare E-discovery: The Light Is Green, But Proceed With Caution*, AHLA HEALTHCARE LIABILITY & LITIGATION, Jan. 13, 2012.

[44] Predictive coding will likely transform the economics of discovery in patent litigation. *See* Daniel B. Garrie, *Change is Coming: The Evolution of E-*

C. The Federal Circuit Should Modify the Model Order to Allow Parties to Work with Predictive Coding and Similar Technologies

Since the Model Order presumptively limits the number of custodians and search terms for all email production requests, the Model Order isolates the e-discovery focus on permissible keywords and custodians, which substantially decreases the benefits afforded by predictive coding.

The Model Order, while useful for limiting keyword search discovery demands, is not framed for other technologies like predictive coding. While predictive coding uses keywords, it does not utilize them in the same manner as keyword-driven technology. Consequently, Model Order's value in limiting keywords and custodians hinders any ability to use predictive coding technologies.

CONCLUSION

Courts and counsel should utilize the Model Order as a starting point for assessing the e-discovery process in patent disputes, but should also consider the potential pitfalls that the Model Order presents. As a few cases have shown since implementing the Model Order, the court is willing, within reason, to allow parties to produce their own mutually agreeable protocol.[45] However, it remains to be seen what will happen in a case involving unwilling parties whose case demands more than what the Model Order allows.

Discovery Economics, THE METROPOLITAN CORPORATE COUNSEL, Nov. 28, 2012, http://www.metrocorpcounsel.com/articles/21299/change-coming-evolution-e-discovery-economics; Corinne L. Giacobbe, Note, *Allocating Discovery Costs in the Computer Age: Deciding Who Should Bear the Costs of Discovery of Electronically Stored Data*, 57 WASH. & LEE L. REV. 257, 282-83 (2000); AAB Joint Venture v. United States, 75 Fed. Cl. 432, 438, 442-43 (2007) (ordering discovery of storage tapes defendants conceded might have relevant information despite $150,000 restoration cost).

[45] *See, e.g.*, McGrath v. United States, No. 11–318C, 2012 WL 726423 (Fed. Cl. March 6, 2012) (modifying a proposed discovery order submitted by the parties that was based, in part, on the Model Order).