Bloomberg BNA

Digital Discovery & e-Evidence®



DATABASES

The authors examine how much—if any—protection from discovery is afforded to databases attorneys create for use in complex litigation.

# Can Lawyers Be Compelled to Produce Data They Compile? An Emerging Front in the Trenches of eDiscovery Battles

By Fernando M. Pinguelo and Mason A. Barney

**Fernando M. Pinguelo**, *a partner at Scarinci Hollenbeck and Chair of its Cyber Security & Data Protection group, is a trial lawyer who devotes his practice to complex business disputes with an emphasis on cyber security, data privacy, media and employment matters in federal and state courts. Pinguelo earned accreditation as an information privacy professional (CIPP/US) from the International Association of Privacy Professionals. He is the author of numerous publications and lectures internationally on law and technology issues, including for the Judicial College and Escola Paulista da Magistratura, São Paulo. He may be reached at fernando@CyberJurist.com.*

**Mason A. Barney** *is an associate at Olshan Frome Wolosky LLP whose practice focuses on complex commercial litigation with a concentration on cyber law and technology issues. To learn more about Barney, visit http://www.olshanlaw.com/attorneys.*

Every week it seems we hear a new story about the critical role that big data and data mining plays in modern business. Many companies keep proprietary databases with thousands of pieces of information about millions of people around the world. With recent news solidifying Amazon's present dominance over cloud computing and its less familiar business that rents processing power to companies as a backdrop, big data analytics is an industry expected to reach $125 billion worldwide this year, and it is only growing.[1]

As with the business community, the power of big data has had and will continue to have a major impact in complex litigation. However, if an attorney were to create a database to assist in understating the mountains of information a large litigation can generate, then would that database be considered protected work product—or is it just another discoverable document that opposing counsel can request?

Those are the key questions raised recently by a motion to compel filed in the District of Massachusetts on April 15, 2015 by the Equal Employment Opportunity Commission (EEOC). The Massachusetts federal district court's recent decision offers a rare glimpse into the brave new world of ''coding'' in the context of litigation strategy and readiness and could have a significant impact on how attorneys approach data acquisition, analysis and coding.

## *EEOC v. Texas Roadhouse, Inc.*

In *EEOC v. Texas Roadhouse Inc.*, D. Mass, 1:11-cv-11732, 5/5/15, the EEOC alleges that Texas Roadhouse, Inc., a nationwide chain of casual dining restaurants,

[1] http://www.forbes.com/sites/gilpress/2014/12/11/6-predictions-for-the-125-billion-big-data-analytics-market-in-2015/

 ISSN 1941-3882

engaged in a pattern and practice of refusing to hire people over the age of 40.

As part of discovery, Texas Roadhouse produced hundreds of thousands of hand-written job applications. To help make sense of the massive volume of information contained in the applications, Texas Roadhouse's counsel had the hand-written applications transcribed into an electronic database. In doing so, counsel separated or ''coded'' the information into data fields such as the applicant's name, address, and social security number. This coding process cost Texas Roadhouse tens of thousands of dollars to implement.

## The Motion to Compel

Texas Roadhouse produced the applications as images to the EEOC, but without the coded information.

In April 2014, the EEOC asked whether Texas Roadhouse was building a searchable database of applications and, if so, whether it would agree to share that database with the EEOC. The EEOC reasoned that any such database represented merely a reformatting of the data that had already been produced, and as such, was responsive to its document requests.

Texas Roadhouse confirmed that it did create a database, but that it would not share it because it represented protected work product.

In November 2014, the EEOC raised the issue of access to this data during a conference with the court. The court did not order production, but suggested that the parties try to negotiate a data sharing arrangement.

When those negotiations failed, the EEOC moved to compel production of a list of the fields used by Texas Roadhouse in processing the database, along with identification of the cost associated with such processing. The EEOC planed to use this information in a second, ''targeted'' motion to compel production of the underlying data in the relevant fields.

## EEOC's Arguments In Favor of Disclosure

In its motion, the EEOC argued that, contrary to Texas Roadhouse's assertion, the data it seeks are not protected by the work-product doctrine. The fact that counsel created the database during the litigation was '' a distinction without a difference,'' according to the EEOC.

In its argument, the EEOC characterized the database as a mere mechanical compilation of already existing facts that is not entitled to work product protection from discovery. It averred that the simple notation of an applicant's name, address, social security number etc. does not reveal the attorney's thought processes or analysis of the case. The EEOC argued further that it had a substantial need for the data to eliminate any later arguments regarding the accuracy of data used by experts, and it noted additionally that the cost to create its own database would impose an undue hardship.

## Cases Supporting Disclosure

In support of its argument, the EEOC relied primarily on two cases. The first, *Portis v. City of Chicago*, 2004 BL 4161 (N.D. Ill. July 7, 2004) was a class action where the defendants moved to compel production of a database plaintiff's counsel created to compile information pulled from more than 20,000 arrest reports of potential class members. The court granted the motion. While the court agreed with plaintiff that the database was work product and that counsel's selection of what information to include could disclose what adverse counsel thought were the important data points in the case, it concluded nonetheless that the sheer size of the dataset ''virtually eliminates the possibility that defendants could discern plaintiffs' litigation strategy from the database.''

Furthermore, the court reasoned that defendants had a substantial need for the database because access to the data would ''materially advance the litigation'' by helping defendants evaluate the proposed class list more efficiently, and the amount of time and money it would take to recreate the database warranted a finding of undue hardship.

The EEOC also relied on Judge Scheindlin's opinion in *S.E.C. v Collins & Aikman Corp.*, 256 F.R.D. 403, 411 (S.D.N.Y. 2009). There, the SEC brought charges of insider trading against several defendants. The defendants asked in their initial document requests that the SEC produce documents that support the allegations in the complaint.

In response, the SEC produced more than 1.7 million documents. However, through correspondence with the SEC the defendants learned that, in preparation for drafting the complaint, the SEC's counsel had segregated many of these documents into 175 folders that correlated to the allegations set forth in the complaint. The defendants argued that by producing millions of documents, without identifying which documents were segregated into those folders, the SEC had engaged in an impermissible ''document dump.'' Consequently, defendants had no choice but to move to compel the SEC to identify the documents in these folders.

**According to the company, the selection of what data to include deserved absolute protection from disclosure under the work-product doctrine.**

The court rejected the SEC's contention that its segregation of these documents was protected by the work-product doctrine. Instead, it ordered the SEC to identify the documents in the folders.

The court reasoned that because Fed. R. Civ. P. 11 requires a party to have support for the allegations in a complaint, ''producing the compilations of documents that support the factual allegations of a complaint reveals no more than that already revealed by the filing of the complaint''; and thus, the segregation of documents could not be considered core work product.

The court then held that even if the segregation were to qualify for work product protection, then defendant had a substantial need to know what documents were in the 175 folders in order to understand what documents supported the allegations against him.

Further, even though the defendant could have sifted through the documents himself, the time and expense needed to do so ''constitutes 'undue hardship' by any definition.''

  DDEE ISSN 1941-3882

## Texas Roadhouse Arguments Against Disclosure

In its opposition, Texas Roadhouse focused on the fact that the database it created did not include all of the information contained in the applications; rather, the database included only that data the attorneys thought were important to support the company's defenses. The process ''requires judgments as to what raw data to use, what raw data not to use, and how to classify the data.'' The company argued that the choices it made at the time it created the database and coding showed what type of statistical analyses it planned to use and thus, revealed the company's trial strategy.

Moreover, if during the inputting process the attorneys decided to change what data they were inputting or how they were categorizing the data, then that would further highlight the company's strategy and may show, among other things, where the attorneys thought strengths or weaknesses existed in their case. As such, according to the company, the selection of what data to include deserved absolute protection from disclosure under the work-product doctrine.

## Federal District Court Sides With Texas Roadhouse

In a perfunctory order entered on May 5, 2015, the magistrate judge agreed with Texas Roadhouse and denied the SEC's motion. The district court found ''persuasive defendants' argument that their decisions regarding what fields to code, and their work coding documents thus far, is work product that is not discoverable.''

## Practical Considerations for Attorneys.

Even though the court agreed with Texas Roadhouse, the holdings in *Portis* and *Collins & Aikman*, along with the arguments raised by the SEC, demonstrate that attorneys must be mindful about the potential that any compilation of facts that they create could be made discoverable. *See also Williams v. E.I. du Pont de Nemours & Co.*, 119 F.R.D. 648, 651 (W.D. Ky. 1987) (ordering production of a database created by defendant's counsel that could potentially be used by both parties' experts); *Fauteck v Montgomery Ward & Co., Inc.*, 91 F.R.D. 393, 399 (N.D. Ill. 1980) (same).

Clearly, as the court found in *Texas Roadhouse*, not every data set created by an attorney is discoverable.

Courts still apply the traditional tests for work-product protection and do not order production where doing so could reveal the attorney's mental impressions. *E.g., In re Bloomfield Mfg. Co.*, 977 S.W.2d 389, 392 (Tex. App. 1998) (holding that a database that included a paralegal's analysis of the relevance of certain accidents was protected work product). But decisions such as *Collins & Aikman* demonstrate that the line between opinion and fact is not always so clear.

**Most courts acknowledge that if a database compiled for litigation were to be produced, then the parties should share in the cost of its production.**

Furthermore, the rulings in *Portis* and *Collins & Aikman* demonstrate that attorneys can not always rely on the arguments made by Texas Roadhouse that their careful selection of what data to include in such data compilations shows their mental impression of the case. As noted, in order to succeed, Texas Roadhouse needed to establish specifically that the selection of data would reveal the types of statistical analyses that it was intending to use, and the analyses it was choosing to not use.

Practitioners should also be aware that most courts acknowledge that if a database compiled for litigation were to be produced, then the parties should share in the cost of its production. *E.g.*, *Portis*, 2004 BL 4161 at *5 (ordering sharing of costs ''[i]n order to avoid seriously prejudicing plaintiffs); *Williams*, 119 F.R.D. at 651.

But this is not universally applied, and even when cost sharing is ordered, that does not mean that the producing party will be made whole. *See* , *Collins & Aikman Corp.*, 256 F.R.D. at 411 (ordering production without mentioning cost sharing); *Hines v Widnall*, 183 F.R.D. 596, 601 (N.D. Fla. 1998) (noting that the ''fair portion'' cost sharing provisions of Rule 26 only applied to expert discovery); *see also Portis v City of Chicago*, 2004 BL 4167 at *5 (N.D. Ill. Dec. 7, 2004) (reducing the amount plaintiff would be reimbursed based on the prevailing hourly rate for paralegals in Chicago).

As Texas Roadhouse argued, it created the applications database to aid in its defense, not to help the EEOC sue it.

However, even though the court agreed with this proposition, the cases noted above demonstrate that this argument alone may not carry the day. The takeaway from these cases is that attorneys must be cognizant that when they create databases in order to analyze big data sets, they may see their hard work used against them.