

May 9, 2013

Court Permits Combination of Predictive Coding and Keyword Search

Focusing on precision rather than recall, district court finds that process complies with discovery obligations.

On April 18, the U.S. District Court for the Northern District of Indiana issued a discovery order in *In re Biomet M2a Magnum Hip Implant Products Liability Litigation*,¹ finding that defendant Biomet's discovery process, which included the combined use of keyword search and predictive coding, fulfilled its discovery obligations. However, the court accepted Biomet's reliance on **precision** measurements, rather than **recall** measurements, leading to a potentially substantial underestimation of what proportion of relevant documents Biomet produced.

Background

In response to the plaintiffs' discovery demands, Biomet collected 6 terabytes of data and filtered the resulting 19.5 million documents with keyword searches to identify approximately 3 million documents for review.² They performed a predictive coding review on these 3 million records to identify documents for production, but the plaintiffs objected to Biomet's approach, arguing that Biomet should have applied predictive coding to all 19.5 million documents and should be required to do so to find any remaining relevant documents. The plaintiffs alleged that the use of keywords **before** applying predictive coding polluted the results of the process. The plaintiffs also argued that Biomet should have allowed the plaintiffs to participate in a joint review of the documents used to train the predictive coding software. Biomet did offer the plaintiffs the opportunity to propose additional keyword searches and invited the plaintiffs to review samples of the output of the predictive coding system.

Court's Opinion and Biomet's Statistical Claim

The court rejected the plaintiffs' arguments, focusing its analysis on whether Biomet had satisfied its obligations under Federal Rules of Civil Procedure 26(b) and 34(b)(2) and the Seventh Circuit Principles Relating to the Discovery of Electronically Stored Information. The court found nothing in the duty of cooperation that requires the parties to jointly review data. It also deflected the plaintiffs' argument that limiting the document population with keywords prior to applying predictive coding necessarily diluted the value of the latter process. The court also focused on the cost of the review of all 19.5 million documents proposed by the plaintiffs, finding that the costs were not proportional to the "comparatively modest" increase in the relevant documents that would be found, as based on the statistical testing performed by Biomet.³

Biomet's brief in support of its process was the source of the statistical claim that only 0.94% of documents not hit by its keyword searches were relevant. Its expert characterized this as a "very low number of potentially responsive documents" missed compared with the 16% relevance of the keyword search results, which the court

1. *In re Biomet M2a Magnum Hip Implant Prods. Liab. Litig.*, No. 3:12-MD-2391 (N.D. Ind. Apr. 18, 2013) (order regarding discovery of ESI), available at <http://www.innd.uscourts.gov/mdl/ESI%20order.pdf>.

2. Biomet also used de-duplication to reduce the number of documents for review.

3. *Biomet* order, *supra* note 1, at 5.

Morgan Lewis

echoed in its order. While, the 0.94% figure is comparatively small when measured against the 16% relevance of the keyword search results, it represents a much larger number of actual documents that the percentages seem to indicate. Biomet's measurement showing 0.94% relevance equates to approximately 86,000–210,000 missed responsive documents. Compared with the approximately 180,000–230,000 relevant documents the keywords did retrieve, the keyword searches potentially excluded more responsive documents than they retrieved.

Implications

Courts continue to issue orders and opinions allowing (and occasionally requiring) the use of predictive coding as a means of reducing the cost of discovery. The court in *Biomet* accepted the notion that predictive coding is a reasonable method by which a party may meet its discovery obligations and that cost shifting can be an appropriate means of addressing proportionality concerns. It made clear that cooperation does not require complying with the requesting party's demand for a specific process, and it was also not convinced that keyword search and predictive coding cannot be used together, as the plaintiffs argued.

It is clear, however, that the court did not base its reasonableness assessment on a measure of the level of **recall**⁴ of Biomet's process. Instead, it focused on comparative costs and Biomet's assertions that the keyword search results had a greater proportion of relevant documents than the documents that were not hit by the keyword searches. This focus on **precision** rather than **recall** led the court to approve Biomet's process, which may well have left behind more relevant documents than it found.

It is critical to remember that the standards for discovery are reasonableness and proportionality, not perfection. 100% recall of relevant documents is not required by courts' rules, but producing parties should not rely solely on the type of comparative **precision** measurements that the court agreed with in *Biomet*. They should instead focus on achieving reasonable **recall** rates while defensibly managing costs and risks given the specifics of each case. Strategies to achieve this may include limiting the scope of collection, applying keyword searches, using predictive coding, and employing other methods depending on the matter.

Contacts

If you have any questions or would like more information on the issues discussed in this LawFlash, please contact any of the following Morgan Lewis eData attorneys and technologists:

Attorneys

Stephanie A. "Tess" Blair	Philadelphia	215.963.5161	sblair@morganlewis.com
Scott A. Milner	Philadelphia	215.963.5016	smilner@morganlewis.com
Jacquelyn A. Caridad	Philadelphia	215.963.5275	icaridad@morganlewis.com
Tara S. Lawler	Philadelphia	215.963.4908	tlawler@morganlewis.com
Denise E. Backhouse	New York	212.309.6364	dbackhouse@morganlewis.com
Lorraine M. Casto	San Francisco	415.442.1216	lcasto@morganlewis.com
Graham B. Rollins	Washington, D.C.	202.739.5865	grollins@morganlewis.com
Jennifer Mott Williams	Houston	713.890.5788	jmwilliams@morganlewis.com

Technologists

L. Keven Hayworth	New York	212.309.6929	khayworth@morganlewis.com
James B. Vinson	Philadelphia	215.963.5391	jvinson@morganlewis.com
Wayne R. Feagley	San Francisco	415.442.1737	wfeagley@morganlewis.com
George E. Phillips	Houston	713.890.5769	george.phillips@morganlewis.com
Jessica A. Robinson	Washington, D.C.	202.739.5784	jessica.robinson@morganlewis.com

4. Recall is the actual proportion of relevant documents retrieved out of a population of documents being searched. A related measure, precision, is the proportion of ultimately relevant documents within a set of documents retrieved by a given search.

Morgan Lewis

About Morgan, Lewis & Bockius LLP

With 24 offices across the United States, Europe, and Asia, Morgan Lewis provides comprehensive litigation, corporate, transactional, regulatory, intellectual property, and labor and employment legal services to clients of all sizes—from globally established industry leaders to just-conceived start-ups. Our international team of lawyers, patent agents, benefits advisers, regulatory scientists, and other specialists—more than 1,600 legal professionals total—serves clients from locations in Almaty, Beijing, Boston, Brussels, Chicago, Dallas, Frankfurt, Harrisburg, Houston, Irvine, London, Los Angeles, Miami, Moscow, New York, Palo Alto, Paris, Philadelphia, Pittsburgh, Princeton, San Francisco, Tokyo, Washington, D.C., and Wilmington. For more information about Morgan Lewis or its practices, please visit us online at www.morganlewis.com.

This LawFlash is provided as a general informational service to clients and friends of Morgan, Lewis & Bockius LLP. It should not be construed as, and does not constitute, legal advice on any specific matter, nor does this message create an attorney-client relationship. These materials may be considered **Attorney Advertising** in some states. Please note that the prior results discussed in the material do not guarantee similar outcomes. Links provided from outside sources are subject to expiration or change. © 2013 Morgan, Lewis & Bockius LLP. All Rights Reserved.