

UNITED STATES DISTRICT COURT  
NORTHERN DISTRICT OF INDIANA  
SOUTH BEND DIVISION

IN RE: BIOMET M2a MAGNUM HIP	)	
IMPLANT PRODUCTS LIABILITY	)	
LITIGATION (MDL 2391)	)	CAUSE NO. 3:12-MD-2391
	)	
	)	
	)	
<hr/>	)	
This Document Relates to All Cases	)	
<hr/>	)	

ORDER REGARDING DISCOVERY OF ESI

Biomet has produced 2.5 million documents to plaintiffs in this docket's constituent cases, and the Plaintiffs' Steering Committee believes production should run to something closer to 10 million documents. The parties have set forth their positions on the procedures or protocols that should be used to facilitate identification, retrieval, and production of electronically stored information in submissions filed on April 1 and 5. The parties seek my guidance as to the direction discovery of ESI should take, and I believe the parties need a prompt ruling more than they need extensive discussion of each point they raise.

Biomet began producing documents in cases eventually centralized here in the summer of 2012. Some plaintiffs' counsel, anticipating this docket's formation, told Biomet (occasionally in forceful terms) not to begin document production until the Judicial Panel on Multidistrict Litigation decided whether to centralize. Biomet, neither sold on centralization nor free of judicial exhortations in other cases against it, started the process of identifying and producing documents.

Biomet used a combination of electronic search functions to identify relevant documents. Keyword culling was used first, reducing the universe of documents and attachments from 19.5 million documents to 3.9 million documents, comprising 1.5 terabytes of data. Removal of duplicates left 2.5 million documents and attachments. Statistical sampling tests of a random sample projected, with a 99 percent confidence rate, that between .55 and 1.33 percent of the unselected documents would be responsive and (with the same confidence level) that between 1.37 and 2.47 percent of the original 19.5 million documents were responsive. In comparison, Biomet's keyword/deduplication approach had identified 16 percent of the original 19.5 million.

Biomet then employed technology-assisted review, or predictive coding, to identify the relevant documents to be produced from the 2.5 million that emerged from the keyword and deduplication processes. Predictive coding has found many uses on the Internet. Under predictive coding, the software "learns" a user's preferences or goals; as it learns, the software identifies with greater accuracy just which items the user wants, whether it be a song, a product, or a search topic. Biomet used a predictive coding service called Axelerate and eight contract attorneys to review a sampling of the 2.5 million documents. After one round of "find more like this" interaction between the attorneys and the software, the contract attorneys (together with other software recommended by Biomet's e-discovery vendor) reviewed documents for relevancy, confidentiality, and privilege.

To date, Biomet's e-discovery costs are about \$1.07 million and will total between \$2 million and \$3.25 million.

Biomet invited the Plaintiffs' Steering Committee to suggest additional search terms and offered to produce the rest of the non-privileged documents from the post-keyword 2.5 million so the Steering Committee can verify that Biomet is producing the relevant documents. The Steering Committee has declined those offers, believing they are too little to assure proper document production.

The Steering Committee contends Biomet's initial use of the keyword approach has tainted the process. They point to a recent article that mentioned unidentified "literature stating that linear review would generate a responsive rate of 60 percent and key word searches only 20 percent, and [the defendants in the case being discussed] proposed that predictive coding at a 75 percent responsive rate would be sufficient." Barry Kazan and David Wilson, TECHNOLOGY-ASSISTED REVIEW IS A PROMISING TOOL FOR DOCUMENT PRODUCTION, New York Law Journal (Mar. 18, 2013). The Steering Committee sees Biomet's approach as insufficient because, although it employed predictive coding, Biomet began with the less accurate keyword search. The Steering Committee sees Biomet's offer to let the Steering Committee propose search terms as unhelpful because the Steering Committee's unfamiliarity with Biomet terminology prevents them from making suggestions advisedly.

The Steering Committee wants Biomet to go back to its 19.5 million documents and employ predictive coding, with plaintiffs and defendants jointly

entering the “find more like this” commands. Biomet objects on a variety of grounds, including its estimate that virtually starting over would cost it millions more than the millions it already has spent in document production. The Steering Committee responds that Biomet gambled when it spent millions on document production that several of plaintiffs’ counsel warned Biomet not to undertake until the Panel had centralized the cases.

The issue before me today isn’t whether predictive coding is a better way of doing things than keyword searching prior to predictive coding. I must decide whether Biomet’s procedure satisfies its discovery obligations and, if so, whether it must also do what the Steering Committee seeks. What Biomet has done complies fully with the requirements of Federal Rules of Civil Procedure 26(b) and 34(b)(2). I don’t see anything inconsistent with the Seventh Circuit Principles Relating to the Discovery of Electronically Stored Information. Principle 1.02 requires cooperation, but I don’t read it as requiring counsel from both sides to sit in adjoining seats while rummaging through millions of files that haven’t been reviewed for confidentiality or privilege. Both sides cite reports from the Sedona Conference project, *e.g.*, The Sedona Conference, *The Sedona Conference Commentary on Proportionality in Electronic Discovery* (Jan. 2013); The Sedona Conference, *The Sedona Conference Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery*, 8 Sedona Conf. J. 189 (2007); and The Sedona Conference, *Conducting E-Discovery After Amendments:*

*The Second Wave*, 10 Sedona Conf. J. 215 (2009), and I don't see Biomet's approach as running afoul of any of the principles set forth in those publications.

In contrast, the Steering Committee's request that Biomet go back to Square One (more accurately Square Two, since Biomet first collected the 19.5 million documents) and institute predictive coding at that earlier stage sits uneasily with the proportionality standard in Rule 26(b)(2)(C). Doing so would entail a cost in the low seven-figures. The confidence tests Biomet ran as part of its process suggest a comparatively modest number of documents would be found. The Steering Committee challenges that conclusion by pointing to studies (one in 1985) indicating that, on average, Boolean searches identify less than a quarter of the relevant documents in a set of documents. Boolean language provides the basis for keyword searches, though I can't find anything in this record that equates today's keyword searches to Boolean searches. In contrast, the Steering Committee says predictive coding identified 75 to 95 percent of the relevant documents — about four times more efficient than keyword searches. The 75 percent figure appears to come from the previously-cited recent New York Law Journal article about technology-assisted review as part of document production. The article itself doesn't vouch for the accuracy of the 75 percent figure; the article simply notes that in a Virginia state court case, Global Aerospace v. Landow Aviation, No. CL 61040 (Va. Cir. Ct., Loudon County, Apr. 23, 2012), the defendants "proposed that predictive coding at a 75 percent responsive rate would be sufficient." I can find no source for the 95 percent figure.

It might well be that predictive coding, instead of a keyword search, at Stage Two of the process would unearth additional relevant documents. But it would cost Biomet a million, or millions, of dollars to test the Steering Committee's theory that predictive coding would produce a significantly greater number of relevant documents. Even in light of the needs of the hundreds of plaintiffs in this case, the very large amount in controversy, the parties' resources, the importance of the issues at stake, and the importance of this discovery in resolving the issues, I can't find that the likely benefits of the discovery proposed by the Steering Committee equals or outweighs its additional burden on, and additional expense to, Biomet. FED. R. CIV. P. 26(b)(2)(C).

The Steering Committee appears to argue that Biomet is estopped from relying on proportionality arguments based on the incremental cost of what the Steering Committee seeks because Biomet embarked on its document identification in disregard of pre-centralization warnings and advice from some counsel for plaintiffs in individual cases. It might be that the Steering Committee's argument could carry the day in some cases, but this one doesn't seem to be such a case. The Steering Committee hasn't argued (and I assume it can't argue) that Biomet had no disclosure or document identification obligation in any of the cases that were awaiting a ruling on (or even the filing of) the centralization petition. Until the MDL Panel enters a centralization order under 28 U.S.C. § 1407 (or transfers a tag along pursuant to an earlier centralization order), a transferee court is free to act on pending matters. Indeed, through its conditional transfer

orders, the Panel regularly encourages transferee courts to do so. To hold that a party that behaves as the transferee court directs, or that follows the transferee court's standing procedures, does so only by forfeiture of the proportionality provision of Rule 26(b)(2)(C), seems an uncongenial exercise of whatever discretion I have. It also would seem inconsistent with the purposes of centralization under § 1407.

In making this ruling, I assume that Biomet will remain open to meeting and conferring on additional reasonably-targeted search terms and to producing the non-privileged documents included in the statistical sample. Beyond that, if the Steering Committee wishes production of documents that can be identified only through re-commenced processing, predictive coding, review, and production, the Steering Committee will have to bear the expense.

ENTERED: April 18, 2013

/s/ Robert L. Miller, Jr.  
Judge, United States District Court  
Northern District of Indiana