dri™ The Voice of the Defense Bar

E-Discovery Connection
The Newsletter of the Electronic Discovery Committee

December 20, 2013 Volume 8 Issue 3

DRI Resources

Featured Articles

Using a Data Analytics-Driven Discovery Workflow

by Amy Fiterman and P. Sean d'Albertis

Introduction

There has been a buzz the last few years about early data assessment ("EDA") or early case assessment ("ECA") (collectively referred to as "EDA" throughout the remainder of this article). When most people think of EDA, however, it is often limited to using a vendor/software tool to cull data early in a case before processing and full review begins. In reality, data assessment, or what may also be called data analytics, can be much more than this and may or may not include the use of vendor tools. Indeed, various types of analytics should inform discovery strategy by creating an iterative process in the discovery workflow. The use of ongoing data analytics—not just early data assessment—further augments traditional models by redefining and recombining many of the activities from the traditional model in circular iterations. The use of ongoing data analytics across the discovery cycle creates value by helping a party understand its data better, which in turn helps it understand its case better.

The Traditional Discovery Workflow

Traditional depictions of the electronic discovery process typically proceed from left to right in a linear progression. These traditional depictions set forth discovery in self-contained stages such as identification, preservation, collection, processing, review and production. *See, e.g.,* Electronic Discovery Reference Model ("EDRM") at www.edrm.net (in addition to providing standard linear discovery workflow, EDRM has also played a role in exploring how the various stages of discovery can be understood in more non-linear, iterative terms). The utility of such conceptual frameworks consist in charting a path that begins broadly with preservation, followed by a narrowing of the overall data set through a tailored collection, the application of culling criteria and search terms, human review, and finally production. Successive stages of analysis result in an ongoing reduction in volume of the overall data set while increasing the concentration of responsive documents.

In E-Discovery Connection

Committee Leadership


Committee Chair
John J. Jablonski
Goldberg Segalla
(716) 566-5400
jjablonski@gildbergsegalla.com


Committee Vice Chair
Thomas M. Jones
Cozen O'Connor
(206) 340-1000
tjones@cozen.com


Publications Chair
Heyward D. Bonyata
Nelson Mullins Riley & Scarborough LLPs
heyward.bonyata@nelsonmullins.com

Click to view entire Leadership

Seminar

The use of EDA reflects a strategy to identify and realize efficiencies in the traditional, linear model. EDA techniques represent an advancement over traditional models by moving search and culling activities upstream to a position that precedes data ingestion into a processing platform. This shift enhances efficiency and contains cost by reducing the total volume of data that progresses to processing. However, the use of EDA alone misses out on what should be a more comprehensive and less linear approach to discovery through the use of broader data analytics.

Data Analytics

The term data analytics should include any activity directed to understand data sources, composition, and/or relevance. So defined, data analytics can encompass most of the stages set forth in traditional models. This is by design. In practice, the artificial or indeterminate application of traditional stages results in definitions collapsing into one another. For example, if a preservation data set is captured, is this defined as part of the preservation stage or the collection stage? If search terms are used to first identify data for collection is this defined as search, collection, identification, or culling? The strategy described in this article eschews static assignments in recognition that traditional stages might be at play in different ways at different times.

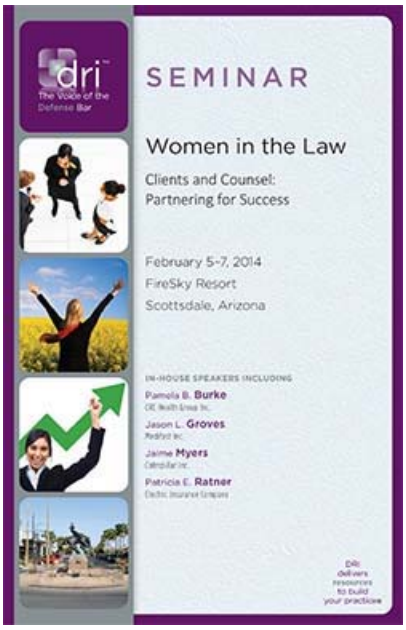
This definition of data analytics is as much about methodology as technology. Understanding data structures and how to locate user created or maintained data sources in a native environment is a basic yet valuable type of data analytics. Such basic scoping might be augmented by applying file extension filters, culling parameters, and search expressions. There are technology solutions on the market that range from a simple utility at nominal cost, to reasonably inexpensive indexing and processing applications, all the way to vastly expensive enterprise deployments that preserve, index, search and collect unstructured data across heterogeneous systems. The key is to start from a methodology designed to understand data and then employ whatever tools are available. It is important to remember that data analytics can be conducted even without expensive specialized tools.

Data Analytics in Action

An iterative feedback loop is the engine that drives the data analytics-driven discovery workflow. By definition, this strategy is flexible and dynamic.

Preliminary Activities. From the moment litigation is foreseeable, a duty to preserve attaches. *See Pension Committee of the University of Montreal Pension Plan v. Banc of America Securities, LLC,* 685 F. Supp. 2d 456 (S.D.N.Y. 2010). Discharging this duty often entails conducting an analysis that integrates both legal and technical expertise. Counsel conducts interviews with client IT, key custodians, and subject matter experts to define the scope of preservation and formulate a preservation strategy.

Scoping and Fact Investigation. Once key custodians are identified, the next step may be to determine whether one or two might be considered representative. Local drive, network shares and e-mail could be culled at the point of collection prior to harvesting. The key concept is to think through opportunities to increase data awareness at each decision



[Women in the Law]

February 5-7, 2014
Scottsdale, Arizona

DRI Publications

DRI Survey of Federal and State Court Electronic Filing Rules

Print to PDF

point rather than applying a prescriptive, data-blind process. The commitment to greet data with inquiry at each step is the cornerstone of a data analytics driven workflow.

Once a sample data set has been harvested, analytics should be applied to cull by objective and subjective criteria, identify content relevant to fact investigation, build a larger custodian list and analyze search term development for discovery responsiveness.

Next Iteration Validation. The benefits of the first step are explored and validated in the next iteration. For example, if the fact investigation part of the analysis reveals avenues to pursue, next steps might involve additional data samples for further evaluation. The analysis from the first step may identify additional custodians as well as create a means to demonstrate that certain custodians are unreasonably duplicative and/or cumulative. Search term development for responsiveness supports the negotiation of an ESI protocol. A party and its counsel should understand the risks of negotiating such terms prior to sampling and testing. Without analysis to support decisions about terms, parties can end up with over-inclusive results that create burdens and costs that overwhelm the amount in controversy. In the alternative, under-inclusive terms may not return probative results.

Successive Iterations. The feedback loop is repeated and refined until the full scope of a project can be defined. Custodian lists, a harvesting plan, and search terms are finalized in a discovery plan. The details of the plan can be stipulated in an ESI protocol/agreement with opposing counsel. The overall project moves downstream as harvested data is processed and loaded into a review platform.

Post-Processing Data Analytics. The focus of data analytics from the beginning of discovery through data processing is to define, analyze, cull, and search the universe of data. Once documents have been loaded into a review database, data analytics can be applied to help understand content and further reduce cost. For example, threading might be used to understand e-mail conversations in context. Batches can be assigned to reviewers in a manner that creates efficiencies. Various platforms can create taxonomies or concept groupings. Finally, predictive coding employs an iterative feedback loop to prioritize documents and refine machine learning with successive coding decisions. Although data analytics in review leverage different techniques than data analytics further upstream, the key to remember is that a flexible, iterative process, whatever that looks like for a given case, rather than one that is "data blind" and rigid, will allow you to better understand your own data and reduce inefficiencies.

Conclusion

Discovery is typically viewed through the prism of cost. The process may focus almost exclusively on responding to requests from opposing counsel. A data analytics-driven workflow affords parties an opportunity to create value by understanding their data and generating a more focused production. This can create efficiencies, contain cost, and support a deeper understanding of the facts and evidence in a given matter.

Amy Fiterman *is a partner in Faegre Baker Daniels LLP's product liability and environmental group. She joined the Firm in 1998 after graduating, with honors, from The George Washington University School of Law. Amy concentrates her practice in the defense of class actions, mass torts, product liability and consumer fraud cases—with a focus on the pharmaceutical and medical device industries. Amy's practice centers on large-scale litigation that involves complicated discovery issues. She has extensive practical experience in leading discovery teams in such matters.*

P. Sean d'Albertis *is the Supervisor of E-Discovery Consulting & Document Review at Faegre Baker Daniels LLP. He graduated from Hamline University School of Law in 1996 and focuses his practice on the identification, preservation and collection of electronically stored information during the discovery process. Sean advises both attorneys and clients on computer forensics, providing both forensic acquisition and consulting services.*

Back...