ALM | LAW.COM

https://www.law.com/thelegalintelligencer/2023/09/06/data-tools-in-litigation-support-miss-crucial-details/

NOT FOR REPRINT

COMMENTARY

Data Tools in Litigation Support Miss Crucial Details

The shift toward "summary" data analytics may present divergent trends: the "overlooking" of potentially relevant data points and the lack of transaction-level, detailed analysis more frequently relied upon in days past.

 September 06, 2023 at 10:47 AM

Litigation

By David H. Glusman, Thomas W. Reinke and Jesse J. LaGrossa | September 06, 2023 at 10:47 AM

Two diametric trends in data analytics have important implications for attorneys involved in commercial litigation:

The development of data analysis tools with the capability to aggregate both structured and unstructured data from a variety of sources allows for the transformation and interpretation of large volumes of data in ways previously unheard of.

However, this shift toward "summary" data analytics may present divergent trends: the "overlooking" of potentially relevant data points and the lack of transaction-level, detailed analysis more frequently relied upon in days past.

As business information technology (IT) systems have evolved, the trend in data analysis has been the aggregation of large volumes of raw data from various systems (i.e., CRM, accounting, timekeeping, etc.) and developing tools to summarize this data as an aid to executive and operational decision-making. The current "buzzword" and hype surrounding artificial intelligence (AI) is the culmination of this trend. This has led to the development of well-known AI and data analysis tools, the prevalence of data warehouses, visual data dashboards, and sophisticated statistical analysis. All these tools provide "summary" level data, which may be useful for decision-makers.

In the IT world, data analysis may encompass four categories: descriptive, diagnostic, predictive, and prescriptive analysis. All of these categories require the aggregation, transformation, and summarization of data. However, in litigation, the "manipulation" of raw data—data that may constitute the contemporaneous recordation of historical events within an entity—could present other challenges. As data is "transformed" from its original state, it could lose its integrity for purposes of litigation—similar to chain of custody issues.

 While these developments may aid company decision-makers with the allocation of resources, they could also be problematic to attorneys who need transaction-level analysis of relevant data fields and the comparison of such data points to original source documents maintained in the ordinary course of business.

Transaction-Level Analysis

We may encounter situations where data fields from a company's databases are "overlooked," either intentionally or unintentionally, by an expert. Similarly, certain fields that theoretically should reside in a company's systems may not be available for extraction—due to the database design or lack of input by the front-end user.

In discovery, companies and their counsel may want to obtain complete, "raw" data sets from the actual software applications used throughout the organization. Likewise, newer data types and sources such as email, transcribed video and audio, and other unstructured data now play a role in litigation. Obtaining a complete set of records maintained in a dataset may provide historical recordation of events central to a dispute. Likewise, the "data dictionary" accompanying the dataset may be insightful for detailed, transaction-level analysis.

The "data dictionary" is a document that describes the database tables, the format and content of individual fields, etc. The data dictionary may be a mechanism for identifying relevant data fields in a commercial dispute. Such a dictionary may include definitions of the following:

- Data Table-A data table is a category of data records; for example, customer demographic information or the fields captured in a sales transaction may constitute a data "table."
- Data Record-A data record captures or depicts events or transactions through various fields. In litigation, data records may constitute a contemporaneous recordation of historical events.
- Data Field-A data field describes the data elements that comprise the data record. For example, an expense disbursement recorded in a company's accounting system may constitute the "data record"—i.e., the transaction-level recordation of an event, but the individual components of the payment (i.e., payment date, amount, expense category, description, etc.) constitute the "data fields" underlying the data record. A data dictionary may provide definitions for each data field, including the original source (i.e., contemporaneous records) of such fields.

The data dictionary is especially important in disputes involving legacy systems—software applications that have evolved over time. In these systems, a core data element may exist in several different tables and records with fields that have different names. These different fields may be populated with different data. Identifying where core data elements or variations exist in a data set is essential. It is also possible several fields with descriptions that appear important may exist but are populated with irrelevant or incorrect data.

Changing Data Analysis Landscape

The wider adoption of statistical computing languages is driving the changing nature of data analysis. A specific example is R™, a software language for statistical computing and graphics. R is a free, open-source software environment that promotes widespread collaboration for creating innovative analytic algorithms. R and multiple other open-source statistical computing languages are appearing in litigation data analysis. The challenge here is that, in many ways, data analysis is becoming a "free-for-all" with longstanding, well-documented, and validated statistical analysis methods giving way to freeform analysis with one-off custom algorithms created for a specific situation. This development creates another set of challenges for litigation support experts.

Going Forward

Data mining has moved away from targeted extraction of data fields. It has become the process of extracting and discovering patterns in large, diverse data sets involving the intersection of machine learning, statistics, and new database structures and looking for patterns, correlations, trends, and anomalies that might be significant. Litigation support experts may need industry-specific knowledge to understand how changes in data mining impact new analytics assertions and outputs that may appear in litigation.

All these developments may have implications for litigation support experts. These experts may need to demonstrate an understanding of new data types, data elements, industry-specific software applications, and other IT developments, including new approaches to data mining and stepped-up adoption of open-source statistical programming software. This may give rise to the need for multiple experts to work parallel—i.e., a data analyst alongside a forensic accountant.

David H. Glusman *is an advisory partner in Marcum LLP's Philadelphia office. He provides consulting services in the areas of forensic accounting, litigation support, health care fraud, and high net worth estates, trusts and taxation. Contact him at david.glusman@marcumllp.com.*

Thomas Reinke *is a manager of health care advisory services at the firm. Contact him at tom.reinke@marcumllp.com.*

Jesse J. LaGrossa *is a manager in the valuation, forensic and litigation support services department at the firm.*

NOT FOR REPRINT