Legal Intelligencer

https://www.law.com/thelegalintelligencer/2022/09/09/litigations-nemesis-dirty-data-presents-complex-challenges/?slreturn=20220812162859

(L-R)David Glusman and Thomas Reinke of Marcum LLP. Courtesy photos

COMMENTARY

Litigation's Nemesis: Dirty Data Presents Complex Challenges

Dirty data often presents complex challenges in litigation cases. The transformation to a digital and cloud-based world is significantly impacting the accuracy and volume of data, as well as the way it is captured and stored. It is also contributing to an increase in "dirty data."

September 09, 2022 at 11:25 AM

By David H. Glusman and Tom Reinke

Dirty data often presents complex challenges in litigation cases. The transformation to a digital and cloud-based world is significantly impacting the accuracy and volume of data, as well as the way it is captured and stored. It is also contributing to an increase in “dirty data.”

- Dirty data can result from any of the following scenarios:
- Intentional fraudulent activities;
- Poor data security;
- Unintentional data entry errors;
- Erroneous reporting;
- Software errors; Transfer of data over periods of time between storage locations or to various software; or
- Poor information systems management.

Cloud-based information systems are becoming more complex and are often widely dispersed to end users, increasing the likelihood of both intentional and unintentional dirty data. Using contract vendors for some routine processes may also increase the likelihood of errors that lead to dirty data.

Dirty data most commonly occurs in transactional information systems such as accounting software, investment/banking/financial systems, billing software, inventory systems, and other applications that record the history of transactions in specific customer, individual or corporate accounts.

The most common litigation cases that arise from transactional systems include various forms of financial fraud, theft and misrepresentation. Litigation may also involve cases that focus on poor performance in management services contracts, such as outsourced medical billing contracts and any transaction with large volumes of activity and data.

The key to investigating and documenting dirty data in litigation cases is getting complete access to the data. When counsel requests or demands data, the opposing party often claims the data is proprietary or the request is unduly burdensome. However, it's important to distinguish between data and the software application data runs on. The application software and other aspects of a data system may be proprietary, but the data and transactions requested likely originated with and are owned by one of the litigants. The counsel may have to make a significant effort for the court to recognize that the data files are essential and that the purpose of the request is not to obtain access to proprietary software or other aspects of the IT system.

Another common response to data requests is to contend that access to the complete transactional database jeopardizes data or system security. One party may instead ask the opposing litigant to specify data and reports to run, or they may provide access to a data warehouse reporting application. Neither option typically helps determine the underlying processes and issues critical to the litigation and dispute. Management reports often do not provide enough detail for the litigants, the attorneys, and the finder of fact, especially when an expert will be involved in reporting to the court on specialized issues. The data warehouse alternative is often inadequate because litigators need access to complete historical files of all transactions and all fields in each transaction, and the primary purpose of a data warehouse is to transform original data fields and transactions into homogenized data sets.

Another important step is to ensure the data is delivered in a usable format. Generally, the standard approach is for the producing party to provide a “dump” of files with data fields and data, often referred to as flat files. However, raw data files are often unreadable and unintelligible in the context of the litigation. It may not be possible to compile these files into a history of transactions. The goal should be to obtain access to standard files that can be read by industry machines and include clearly defined transactions and fields, which may also allow for artificial intelligence to be used in the forensic analysis process.

Litigants also need a data dictionary with all data files. The data dictionary contains written specifications for the data tables, the linkages between tables, the layout of transactions or records, and the definitions of the source and data in each field. It is an essential tool to understand and analyze files, records, transactions and fields.

Other key elements in a data request include control totals, which provide statistics and details on the size of the database, number of tables, number of transactions, and count and amount of totals for key fields. These may also be referred to as “hash totals.”

A digital forensics expert plays a key role in ensuring that data requests are complete and accurate. The primary qualification for a digital expert is firsthand experience in the industry that is the subject of the litigation, or those with similar data. They also need firsthand knowledge of the data sets and transactions, down to the field level, and an understanding of how fields are processed and reported by information systems. Knowledge and experience with the errors, fraud, and misrepresentation that can occur, and how those items appear in data and transactions, is also essential.

A digital forensics expert can play an important role in preparing deposition questions and determining damages. Key capabilities of a digital expert includes understanding the data dictionary. Data dictionaries may need to be expanded and corrected. In addition, the data request should include control totals to define the files provided by the opposition to define the total database.

Any time you receive voluminous data, verifying its integrity and spotting extraneous or erroneous data within the sets is key to obtaining the best answers in the shortest amount of time.

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.*