

INVESTIGATIONS & COMPUTER FORENSICS

Web address: <http://www.nylj.com>

MONDAY, MAY 12, 2008

What a **Difference** a **Database** Makes

Forensic analysis uncovers key information.

BY ALON ISRAELY

DATABASE FORENSICS involves the analysis of data from within a system that contains its information in a database structure, and is used to assist parties to prove or defend a legal position. Database forensics requires technical and analytical skills and can be critical to reliable fact finding and to forming sound legal arguments. It opens a world of data-related possibilities that does not usually exist in other, more common data sets such as e-mail and user-created documents.

Under the Federal Rules of Civil Procedures (FRCP), information contained within a database is Electronically Stored Information (ESI), although ESI is generally described by lawyers and the courts only in terms of e-mail and "e-docs" (e.g., word processing documents, spreadsheets and presentation programs).

E-mail and e-docs are forms of ESI commonly referred to as "unstructured ESI" because of how they are normally maintained: whimsically and discretely, jumbles of information strewn across laptops, file servers and handheld devices.¹ On the other hand, database information is "structured ESI" because it is maintained in an organized manner so that information is sorted, searched and presented logically and powerfully. It is presented so powerfully, in fact, that all major computerized systems rely in large part on some structured set of data (i.e., a database), from using a speed pass at the gas pump to routing air traffic and researching international criminals with Interpol.

Alon Israely is a senior digital evidence expert with BIA and can be reached at Alon@biaprotect.com.

Defining the Term

Databases are collections of fielded (structured) information.²

In relational databases (most common in business), data is stored as discrete sets of information. Those discrete sets are called records and can be thought of as individual documents.

Records are composed of fields that contain values, and those values make up the information content of each record in a database (like sections or chapters in a document).

So for example, in a database that tracks banking transactions, a record may contain fields of information about each individual banking transaction such as transaction number, transaction type, transaction date and amount. It also may contain fields of information about the name, address and social security number of the account holder and fields of information about that particular bank's routing number and branch location from where the transaction was initiated.

Corporations of every size use and rely on databases to manage data related to business operations. Document management systems are one such database; accounting systems, inventory tracking systems, drug trial databases and customer support systems are other examples of databases used in business today.

Some database systems can be complicated and large, spanning many IT resources and used by employees and other systems (integrated databases) across the corporate enterprise. Those systems contain important information about how a company conducts its business and can be critical sources of information for parties involved in litigation or potential litigation.

Litigation Information Sources

Identifying which systems (and therefore, which underlying databases) are relevant to the inquiries at hand is the first step in using databases as a source of information in preparing for litigation. Thus, in a product liability suit the HR diversity tracking database will most likely not be relevant and therefore not be identified as a database to consider.

Once the potentially relevant databases have been identified, a plan must be created in order to get a better understanding of the subset of records that may be relevant to the issues in a particular case. In-house attorneys and the company's outside counsel should speak to the appropriate system administrators (usually via IT) and the business unit managers who own a potentially relevant system in order to facilitate this critical first step. Any strategy that includes mining information within a database to help support a particular legal position requires understanding fielded information contained within the database.

In many cases, information retrieved from databases has helped stopped opposing parties in their tracks with respect to fishing expeditions and other discovery requests. At the same time, analyzing a database has been used to level the playing field by helping counsel uncover a greater universe of facts.

For example, about 10 years ago when e-discovery was still a nascent term and ESI was called "data" or "digital evidence," a small group of attorneys successfully used information from a stock trading clearinghouse database to expand the fact universe which led to a substantial monetary recovery for several clients who had lost their life savings to fraudulent activities by

unscrupulous and criminal stock brokers. As many clients were losing money, the brokers were making money (in private accounts and through exorbitant commissions passed off as other necessary fees).

The attorneys and the teams of experts analyzed the case by poring over thousands of pages of account statements, trading confirmation receipts and other documents. During that process, it became apparent that there were many facts missing, information needed to give a clearer picture of what had occurred during those fraudulent times.

Modern stock trading systems rely on databases. Those statements and trading confirmations were generated by those same trading systems and so by having an opportunity to peer into the guts of those trading systems, the attorneys hoped to uncover more information than existed on the face of those documents already in their possession.

Motion practice began in earnest and fervent arguments and technical discussions ensued in order to determine how best to identify the relevant data. Finally, data was exported from those stock trading systems and formatted as large structured text files and produced to the attorneys for analysis and review.

A team of technologists including database experts, data analysts and programmers assembled to work out the complexities of getting the data to a useful state. In the beginning there were data validation problems which were not discovered, requiring new data exports from the producing party.

For example, one such data validation problem involved the automatic truncation of certain transaction amounts during the export process. Certain fields with money values were cut off during export: A transaction record that may have had a true transaction "buy" amount of 3.12 (i.e., three dollars and 12 cents) had instead been incorrectly exported as 3.1 (i.e., rounded as three dollars and 10 cents).

Thus, data validation is critical to database forensics and analysis. The accuracy of the data is paramount to trusting the subsequent analysis and reporting, so the re-exports without the truncation issue were performed. Once the newly exported data set was validated, the analytics could begin.

Once the analysis was performed, it was determined that much of the information related to each stock trade did not appear on the paper trade confirmation receipts originally provided to the client. That additional information, found only by analyzing the database information, was critical to proving patterns of fraudulent buying, selling and commission distribution.

Defending Against Production

In another example where database forensics techniques were successfully used by parties to prove a particular legal position, a large corporate party was able to prevent unnecessary production of large amounts of data by analyzing the information in a particular relevant database as part of its research for defending itself against burdensome productions. It was a case in which a large financial client had been sued, and one of the parties requested data from all of the systems that tracked a certain type of business transaction at issue in the case.

A team of database analytics experts, system administrators and attorneys familiar with the issues in the case were brought together to make determinations

about whether information from one particular set of databases was indeed relevant to pending document requests. Those databases were analyzed by sampling certain fielded information. Though at first glance the databases seemed pertinent to the inquiries at hand, once the information was analyzed in detail, decisions about relevance that were supported by legal positions were made. Close to 100GB of data was analyzed and determined to be not relevant and therefore, hundreds of thousands of dollars of review costs were prevented.

Thus, regardless of the reasons for instituting database forensics methods for analyzing data, database forensics can be helpful for assisting parties to more properly position themselves for success in litigation.

The analysis of data from
within a system that contains
its information in a database
structure opens a world
of data-related possibilities that
does not usually exist
in other, more common
data sets such as e-mail
and user-created documents.

Important Tips, Considerations

Database forensics requires that data experts and attorneys work together to formulate the correct steps to carefully analyze information contained within databases in order to draw accurate conclusions. Some important considerations include the following.

- *Create a list of potentially relevant systems.* Regardless of whether responding to or submitting a production request, that list will help focus the systems that will require technical and legal concentration.

To create such a list, decide upon the categories of systems that may be relevant. For example, a general list of corporate database systems may include accounting applications, marketing databases, customer portals and technical support ticket systems but only the latter may be relevant if the issues relate to breaches of service levels for an Internet hosting company. Thus, that list should include categories of systems and not necessarily the specific systems themselves (unless known).

- *Consult system administrators or advisers with technical knowledge of potentially relevant systems about the technical nuances of accessing and using the underlying databases.* Sometimes, though an attorney may have a certain argument in mind, accessing the underlying data to support that argument may be impossible or impractical. Those technical advisers are responsible for guiding the attorney down the correct path with respect to the burdens and technical hurdles natural to the system in question.

- *Obtain and use a database mapping.* Database mappings document the fields and structure within a database and also clarify specific values contained

within the fields. For example, a database mapping document may define what the column headings mean, such as "AMT" means "Transaction Amount" and "LOC" means "Location." The mapping document may also define values such as the value, "001" contained in the "LOC" field which stands for "California"—that is, 001 equals the location, "California."

- *Know when to stop.* Sometimes a database will not contain relevant data, and continuing to attempt to extract information for use in litigation may result in unnecessary expenses and lost time. Having an understanding of the limitations of the database, such as the following, are all important considerations:

Can information be extracted efficiently and accurately?

Are there other sources from which the data can be accessed?

What is the likelihood that the data will be relevant to the inquiries?

- *For in-house attorneys, proactive measures can help.* As a proactive measure, in-house legal departments should work, early on with IT and other systems administrators to ensure that current and new systems conform to basic requirements related to data access and use as that may affect future legal situations, and most specifically, discovery obligations. It is better for the legal department in a corporation to work hand-in-hand with IT and other technical departments *before* systems are put in place. This way, once a matter has started, it will not turn out that such systems make it difficult and costly to respond to discovery requests.

Recently, a large corporation organized an internal task force that was responsible for vetting new IT systems with respect to discovery issues. Just as IT security and application engineering groups vet new systems to ensure that those systems integrate appropriately with current business operations, so did this task force work to ensure that new systems (such as e-mail archiving systems and a new web-based accounting module) would not pose undue burdens on the corporation when the data contained within those systems was implicated in lawsuits.

In all, database forensics can help parties uncover facts that will help to drive a successful strategy in litigation and during the research and discovery phases of a lawsuit. E-mail and e-docs are not the only data categories to consider; databases can be just as helpful, and in some cases, more important.

.....●●.....

1. In some cases, ESI is stored in document management systems that help track the use of ESI, such as document versions and e-mail discussions. Document management programs usually store ESI in databases, that is, in some structured system.

2. For technical definitions of databases, see <http://www.lib.fsu.edu/index.php?q=help/libraryterms#D>, <http://en.wikipedia.org/wiki/Database> and the search engine at <http://www.dtic.mil/>.