TECHNOLOGY: WHITE PAPER

ARE YOU MISSING 69 PERCENT OF YOUR INFORMATION DURING DISCOVERY?

By Rob Brunner and Erik Post

THE ISSUES BEHIND THE HEADLINES

ELECTRONIC DISCOVERY OF DATA FROM FINANCIAL AND ENTERPRISE SYSTEMS IS OFTEN NEGLECTED BY ATTORNEYS AWASH IN THE CHALLENGES OF TRADITIONAL E-DISCOVERY PROCESSES

In the past few years, the topic of electronically stored information, or ESI, has become red hot in the legal community. Most of the attention to date has centered primarily on email and electronic documents—documents that fall into the general category of unstructured data—and how to deal with such data during the electronic discovery phase of a lawsuit or regulatory investigation. A variety of processes and tools have emerged in recent years to help companies sift through their massive stores of unstructured data to identify and produce potentially relevant documents for important cases.

However, attorneys increasingly have begun to recognize the important role that structured data—the information that resides in companies' core enterprise systems—can play in prosecuting or defending a case. Indeed, the new Federal Rules for Civil Procedure (FRCP) are clear that the definition of ESI extends well beyond email and text messages to virtually anything stored in a company's systems—including financial statements, sales transactions, and employee payroll and attendance records. In other words, there's no longer any form of digital information that's off limits to a discovery request.

Some definitions may be helpful at this point. Unstructured data comprises documents of mostly natural language text, including word-processing files, e-mail, messaging streams, and text fields from databases. Some documents may include light metadata, such as spreadsheets and RSS feeds. Structured data generally refers to data that is tabular in nature that can be used in computations. The best examples of such data are simple tabular data structures as seen in relational database management systems (RDBMSs), table dumps, and flat files in record format. Semi-structured data generally refers to data documents exchanged between organizations, which often combine unstructured and structured data or (when expressed in XML) text that's structured with metadata tags. Semi-structured data documents of this type usually comply with open standards for data exchange, like SWIFT, NACHA, HIPAA, and EDI.

The FRCP directive is significant given how extensive companies' enterprise systems are and the sheer volume of financial, operational and transaction data coursing through them. For example, a typical Oracle financial database has thousands of tables that can include several terabytes of data. A single year's worth of employee payroll in a large company can encompass millions of records. And point-of-sale systems at mega retailers such as Wal-Mart, Sears or Home Depot can include hundreds of millions of transactions.

In fact, a survey of 370 people by The Data Warehousing Institute found that the volume of structured data far exceeds that of unstructured data. On average, survey respondents estimated that approximately 77 percent of the data in their company's

data warehouse, and 47 percent across their entire organization, is structured in nature. This compares with 9 percent and 31 percent, respectively, of unstructured data. The remaining data in each case, 14 percent and 22 percent, respectively, were described as semi-structured.

The results of The Data Warehousing Institute survey show that while the need for individual attorney review of structured data is not as high as for unstructured data, the relevance, producibility, and need for safeguard and repository of structured data is a multiple of that needed for traditional electronic discovery sources.

STRUCTURED DATA REQUIRES SPECIAL HANDLING

By definition, structured databases—such as those typically found in financial and enterprise IT systems—include fundamentally different types of data and, therefore, require a very different approach to collection, review, analysis and production from what is typically followed in traditional electronic discovery of e-mail files and other unstructured data. Some of these distinctions include the following:

- Unlike unstructured data, which involves "documents" that one can review, structured data includes individual fields, tables and databases containing millions of records.
- While unstructured data can be reviewed and understood on its own, structured data itself may be incomprehensible apart from its host system.
- Often necessary structured data is archived in report format and, thus, must be "reverse engineered" to make analysis possible.
- Because of the size and complexity of some systems, with some structured databases containing thousands of tables, it may be impractical or impossible to include entire systems in an electronic discovery effort.
- "Redaction" in electronic discovery of structured data means masking meaningful information such as Social Security Numbers that may be necessary to relate records together from different data sources.
- Graham-Leach-Bliley, HIPAA and other regulations impose restrictions on the disclosure of customer, employee and patient data, making electronic discovery of structured data even more complicated.

Unlike e-mail, which has a fairly limited number of formats depending on the host system, there are massive differences between the structure and content of data extracted from different types of financial and enterprise systems. Furthermore, even within a type of application—for example, general ledger—significant differences exist from one implementation to the next, as some companies run "off the shelf" versions of popular databases (such as Oracle or SAP) while others run highly customized and modified versions of these same packages and still others use "home grown" applications developed in house.

But despite the challenges involved in electronic discovery of structured data, it can't be ignored or taken lightly. Indeed, this financial and operational data is often critical to the successful prosecution or defense of a case. Consider:

- The employer that saved hundreds of millions by collecting and analyzing hundreds of millions of time/attendance and payroll records.
- The credit card company that successfully defended its turf in an anti-trust investigation by demonstrating the distribution of its customer's purchase patterns.
- The international pharmaceutical company that successfully defended its sales practices by producing individual sales representative call detail records.

— The retailer that defended compliance with a key contract by downloading and analyzing over 250 million sales transactions.

— The bank that proved that its service charges were appropriate based on customers checking account transactions.

— The manufacturer that resolved an FCPA investigation by thoroughly analyzing its global accounts payable information.

Had these companies not been able to access, sort through and produce this critical transactional data as they did, they would have been at risk of losing billions.

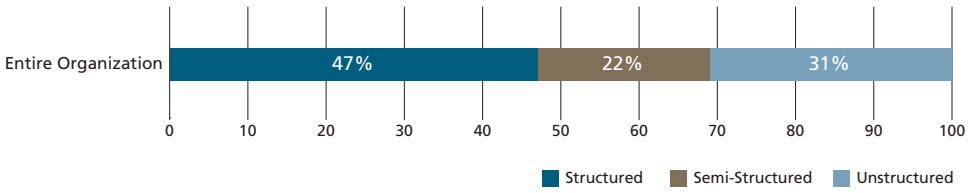
CONCLUSION

To be sure, electronic discovery involving structured data is not easy. Managing the production of financial and enterprise databases in different cases and venues requires tracking at the field, table and database-level. And, with some databases containing thousands of tables, extracting and transforming needed tables and fields is critical. Furthermore, because structured data involves numbers, one can't use keyword searches to help locate potentially relevant data—as is done during discovery of unstructured data, which usually involves words.

As a result, to do electronic discovery of structured data right, companies need robust database analysis tools, as well as professionals with deep functional experience (especially accounting and finance) and expertise in enterprise information management systems and database technologies.

But given the new FRCP rules, companies no longer have an excuse for excluding their operational, transactional and financial data from discovery. And as Amazon discovered, having the ability to identify relevant information hidden deep in the hundreds of terabytes of structured data in enterprise systems can mean the difference between winning big or losing big.

Unstructured and structured data in organizations today



See Philip Russom, "BI Search and Text Analytics: New Additions to the BI Technology Stack," a TDWI Best Practices Report, downloadable at www.tdwi.org/research/reportseries

Rob Brunner and Erik Post are senior managing directors and co-leads of the Financial and Enterprise Data Analytics practice at FTI Consulting, Inc.

FOR MORE INFORMATION CONTACT:

Rob Brunner
Senior Managing Director
rob.brunner@fticonsulting.com
+1 415.283.4204

Erik Post
Senior Managing Director
erik.post@fticonsulting.com
+1 312.252.9338

F T I®

ABOUT FTI TECHNOLOGY

FTI's global team of computer forensics, legal and technology experts help corporations, law firms and governments manage the risk brought by the information age. From data security planning to full-lifecycle electronic discovery software, services and consulting, FTI helps take the time, expense and risk out of managing electronically stored information.