Data Mining: Discovery of Computer-Based Data in Commercial Litigation

Peter Schulman
Denver, Colorado

In commercial cases, valuable information frequently exists on computers. Data mining can be a powerful discovery tool for attorneys and should be considered in the early stages of litigation to minimize costs and avoid the potential loss of data. This author outlines some of the types of information attorneys typically seek and the sources of that information and gives an overview of the data-mining process.

Commercial litigators are frequently confronted with the challenge of identifying, gathering, organizing, analyzing, and presenting voluminous amounts of data. This article focuses on the discovery of computer-based, quantitative data in complex commercial litigation (including commercial bankruptcies). Although this article does not deal with the discovery of nonquantitative data (for example, correspondence and e-mail), some of the principles discussed may apply, particularly with regard to damaged computer media such as so-called "bad" hard drives and unreadable backups.

Frequently, the data sought by attorneys and their experts is on computer media including hard drives, tapes, diskettes, and CD-ROM. As computers and computer software continue to evolve, the existence of information on computer media will increase. The potential types of information sought are almost limitless, but examples include:

- product pricing and cost information
- stock market information
- billing information
- paid vendor history (particularly in commercial bankruptcies)
- revenue and expense information
- financial statements.

The potential sources of information also vary widely. Examples include:

- plaintiffs' and defendants' records

The Trial Lawyer, Vol. 23, 46–48 (2000)

Peter Schulman, CPA, MAE, *is a certified public accountant with a subspecialty in computer consulting. Mr. Schulman has consulted in a wide variety of cases including those involving commercial damages, commercial bankruptcy, anti-trust litigation, and computer fraud. Mr. Schulman has been a practicing CPA for more than 30 years and has acted as a testifying expert for the past 20 years.*

- debtors' records
- public domain (primarily the Internet)
- on-line, fee-based services.

Most attorneys and the experts and consultants who provide professional services to law firms significantly underutilize technology.

THE FUTURE ... IS TODAY!

The legal profession is increasingly computer driven. Many would say that the introduction and evolution of the personal computer (PC) since the early 1980s created a paradigm shift in the legal profession and in many of the professions that provide consulting services to attorneys. When attorneys think about automation, they typically think about word-processing, research, case-management, conflicts-checking, and time-and-billing software. These, and most of the other applications attorneys often think about when they think of computer technology, are word-processing applications, as opposed to data-processing applications. As a result, most attorneys and the experts and consultants who provide professional services to law firms significantly underutilize technology.

Some brief history is interesting as a measure of the extent to which U.S. business has been impacted by the rapid evolution of computer technology:[1]

- In 1970, today's computers would have cost 283 times as much as they cost today. In 1980, they would have cost 22 times as much. In 1990, they would have cost a little more than 3 times as much.
- Spending on computers and peripherals accounted for 19.4 percent of capital spending in early 1997, up from 5 percent at the end of the 1980s.
- Every year, 20 to 35 percent of installed computers have to be replaced because of obsolescence.

THE "BIG PICTURE"

Most data, in its "native"[2] form, is not in a format that lends itself to analysis for litigation purposes. For example, data in an accounting system is often stored in such a way that it can be read and processed only by the accounting system that created it. The process of extracting data (sometimes referred to as "importing" or "exporting" data) so it can be efficiently analyzed and graphed, is sometimes referred to as "data conversion." Usually, the data is analyzed using mainstream database programs such as Microsoft *Access*, *FoxPro*, or *Excel*.[3] Conceptually, data conversion is simple and can be summarized in three steps, as follows:

- Extract the data from the "native" computer file(s).
- Convert the data to a mainstream database format.
- Analyze and, if appropriate, graph the data.

Although the concepts underlying data conversion are conceptually simple, the specific processes involved can be quite complex. Realistically, data conversion is generally not cost-effective for small cases. Fees can vary widely depending upon the facts and circumstances, but they generally start at $10,000 and can be significantly higher. In most litigation applications, primarily because of the contentious environment and because the data set(s) is (are) often unfamiliar, it is difficult to accurately estimate the cost of data conversion projects in advance. As a result, data conversion is developmental in nature; much of the needed information about the data is unknown at the inception of the project and the process is usually laden with unexpected challenges. Challenges such as these are common in other aspects of litigation, so these problems should not discourage attorneys from taking advantage of the benefits of data mining, which can be significant. However, it is important for attorneys and their clients to have a realistic set of expectations to avoid unnecessary disappointments and anxiety.

WHAT ATTORNEYS NEED TO KNOW

Attorneys should be familiar with several concepts to increase the likelihood that computer-based data will be successfully "mined" and to minimize the costs of doing so.

Most data-processing environments are dynamic. In normal business environments, data is routinely added, deleted, and replaced. Under the stresses of litigation, nonroutine processes may also occur that could result in critical losses of data.

For example, data may be intentionally deleted, hard drives may be vandalized,[4] and potentially valuable computer media such as backup tapes and diskettes may be disposed of or overwritten. For these and other reasons, the single most important thing attorneys should keep in mind is that time is of the essence. As a result, if there is any chance of recovering computer-based data in a case, expert assistance should be sought as early as possible. Delays can increase costs and, in some cases, can be fatal due to a permanent loss of data.

The first step in mining data is to identify the type(s) of data to be obtained and the physical location(s) of the data, including backups. These issues are usually addressed in document requests, requests for admission, depositions, and on-site visits. It is important to recognize that successful data recovery requires great attention to technical details. As a result, it is important that attorneys maintain adequate expert involvement.

The single most important thing attorneys should keep in mind is that time is of the essence.

As a practical matter, when it comes to working with experts on technical cases, not all attorneys are "created equal." Many attorneys are not detail oriented and/or they can be impatient when it comes to highly technical cases. As a result, careful attention should be given to the selection of the attorney(s) to work with experts on data-recovery projects. Attorneys with technical educational backgrounds such as mathematics, science, and accounting are frequently the best choice for working with experts on data-mining projects.

Chapter 7 bankruptcy cases usually provide some unique problems. In Chapter 7 cases, it is imperative that computer equipment not be sold or otherwise disposed of until the computers and other important peripherals have been identified. In network environments, file servers must be preserved and, in most cases, at least one or more work stations, including all backup systems, must be preserved.[5]

In cases that involve so called "bad" hard drives and unreadable backups, although the actual methods involved in recovering such data are beyond the scope of this article, attorneys should be aware that there are highly sophisticated resources available to recover data that would otherwise be considered lost forever.

A POWERFUL TOOL

In commercial cases, valuable information frequently exists on computers. Data mining can be a powerful discovery tool for attorneys and should be considered in the early stages of litigation to minimize costs and avoid the potential loss of data.

ENDNOTES

[1]*Forecast: Long-Term Growth,* PC Magazine, August 1997, at 10.

[2]The term *native,* in this context, refers to data in its original form.

[3]Most lawyers and their experts commonly think of Microsoft *Excel* as a spreadsheet program. In addition to being a spreadsheet program, it is also a database program. Although the intricacies of database programs are beyond the scope of this article, there are two general types of database programs, relational databases and flat-file databases. Microsoft *Access* and *FoxPro* are examples of relational databases, and Microsoft *Excel* is an example of a flat-file database. Relational databases are not necessarily better than flat-file databases, and vice versa. They both have their strengths and weaknesses, and the choice of one over the other should be based on the facts and circumstances in each case. In fact, they are not mutually exclusive and can frequently be used together to take advantage of the features of each, particularly in the types of unique applications that frequently present themselves in litigation.

[4]I recently worked on a case where a tiny wire was broken off a debtor's hard drive controller board. The problem could be detected only after close inspection with a magnification device. The hard drive was recovered by replacing the controller board and transferring the data to new computer media which, in this particular case, included copying the "master" to a hard drive and then making copies for use by the consultants and others on CD-ROM.

[5]Backup tape drives sometimes have heads that become misaligned. As a result, it is important to maintain custody of all tape drives, and the computers in which they are installed, so the backup tapes can be read by the tape drives that made them.