

## TWO VIEWS FROM THE DATA MOUNTAIN

By: Steven C. Bennett and  
Thomas M. Niccum, Ph.D.\*

### Introduction

In 1975, when we graduated from high school together, the modern computer age was still in its infancy. Although most major businesses and institutions had some computerized records and operations, the volume of electronic (versus paper) records was still relatively low. The personal computer revolution, desktop networking, the Internet, and e-mail as a common form of business communication all had yet to occur.

These developments, over the last quarter century, for most businesses and institutions have produced a vast mountain of data in electronic form. Many of the most recent developments in computer science and technology, moreover, have made it even easier to store (and, increasingly, to search) this enormous quantity of data.

The ability to create, maintain and use this huge volume of data raises important technical and legal issues. In essence, for most businesses, it is technically possible to keep virtually every electronic record that comes into existence. Indeed, there are costs and other burdens associated with attempting to eliminate electronic records on a selective basis. In many instances, moreover, there may be legal constraints on attempts to destroy electronic records.

We come at this problem from a common background, but from two different professional perspectives. One of us is a computer science professional who largely views the

---

\* Steven C. Bennett is a partner in the New York City offices of the law firm of Jones, Day, Reavis & Pogue and Chair of the firm's E-Discovery Committee. Thomas M. Niccum, Ph.D., is the President of Lancet Software, Inc. in Minneapolis, and **an Adjunct** Professor of Computer Science at the University of Minnesota. The views expressed are solely those of the authors and should not be attributed to the authors' firms or their clients.

data mountain from the perspective of the possibilities of improving the efficiency and productivity of business through effective data analysis and storage. The other is a lawyer who largely views the data mountain with trepidation, knowing that what is buried in the mountain may often be the stuff of which litigation nightmares are made.

Can these two views be reconciled? Although there is no one perfect solution to this problem, we believe that businesses and institutions can, with forethought and sufficient effort, master the basic challenge of the data mountain. The conclusion of this Article is an attempt to outline the most important steps involved in that process.

### Technical Developments And Issues

The data that corporations and other institutions store is growing exponentially. Even small organizations may have hundreds of gigabytes of data stored and available almost immediately, not to mention the backups archived and locked away in off-site storage. Many office workers and professionals have thousands of e-mails (sometimes sorted into folders and sometimes merely kept as a perpetual inbox), recording every scrap of e-conversation. .

As data storage manufacturers continually increase storage capacities and cut costs, our electronic file cabinets are fast approaching a capacity that is effectively infinite. The reasons for this ever-escalating volume of data are many. Nearly every business larger than a paper route uses computers as a normal part of daily operation. Computer systems enable collection of data about sales, inventory, financials and other aspects of business. More and more of this data is retained as firms learn to leverage their investment in information by extracting business trends and customer tendencies from vast warehouses of archived data. This tendency to retain data is accelerated by the decreasing cost of storage. Further, more and more interpersonal and inter-company communication is done electronically by exchanging word processing documents and e-mails. This communication medium is responsible for massive multiplication of documents, as

attachments are added to emails, documents are mailed to multiple recipients and long conversations carried out in e-mail “chains” are copied and responded to over time. All of these electronic items are also being retained for long periods as storage capacities climb.

Advances in data creation and storage are not the only reason that businesses are retaining more and more data. Highly efficient search techniques have also made it possible to use vast quantities of data effectively. Increasingly, due to technology like the systems Lancet has built, it is becoming possible to “mine” these enormous quantities of material. In the past, a company might say, in response to a request by business people (internal) or lawyers (outside, in a lawsuit), “this is the best we can do, given cost and time,” and the results would be limited. Now (and in the future) with effective data mining, it has become possible to pull a lot of information together for a lot less cost.

The same techniques that allow a computer user to search the web for a new broccoli recipe that contains garlic can allow huge numbers of documents to be loaded into a database and searched with sophisticated queries. Fuzzy logic and artificial intelligence techniques allow searches to find and rank documents containing words that are near each other, to find documents that contain some words but not others, or to construct a web of linkages between related documents that allow a user to navigate through the pile in a rational manner.

The implications are significant. What would have looked like a daunting, seemingly impossible research task 25 years ago is now quite possible, and will soon become the norm. Computer users are increasingly aware that, with these sophisticated computerized search mechanisms, millions of records can be reviewed and analyzed, and records in disparate locations can be collected and compared. The data mountain is no longer an impossible height to scale, but a vast database to be mined for secrets and insights that were previously unavailable.

Coupled with the vast expansion of electronic data, and the sharp increase in ability to search and use such data, is the fact that, in the modern computer environment, data tends to persist, often well beyond its intended useful life. Even if an institution has a document retention policy (or, more appropriately named, a document deletion policy), and employees apply the policy correctly by doing their housekeeping (deleting old e-mails and ridding disks of documents) making the data truly disappear is not quite that easy. So-called “deleted data” can continue to exist nearly forever in forms that range from immediately available to quite costly to recover, but the data is recoverable nevertheless.

Discarded data can lurk in a number of spots that the average user may not even know about. Using the ubiquitous personal computer running Microsoft Windows, computer professionals can find data in a wide array of places:

- By default, for most computer users, deleting a file does not truly delete it – (the direction to delete simply moves the file into a special folder called the “Recycle Bin.” And just like a real trash can, if you accidentally toss something in the Recycle Bin, you can retrieve it. For e-mail, there is typically a similar mechanism – deleted items are not deleted, they are just moved into a special folder. Anything in these special folders is recoverable by moving the document back to one of the user’s normal folders.

- Temporary copies of documents are often created during word processing sessions to archive a document in progress in case the computer crashes. This means that there is a “shadow” version of the document, stored on the computer’s hard drive, in a place that the user does not know about – but that document recovery experts DO know about. There are many other places where such “shadow documents” could be hiding., waiting to be recovered by computer forensics experts.

- While a user can “empty” the Recycle Bin – the document is then no longer easily recoverable – at least by that user. But the electronic data persists. A simple comparison to the physical world can help illustrate this problem. When we were children, libraries still had card catalogs. If a librarian were told to remove a book from the library collection, the librarian would go to the card catalog, find out where the book was and then physically remove the book and the card. A computer appears to do the same thing, but accomplishes the removal task in a different manner. A computer, like a card catalog, also keeps track of shelf space, with a map that indicates which slots on shelves are occupied and which are empty. If we tell a computer to “really” delete a document (rather than merely to move it to the Recycle Bin) the actions taken are similar to the librarian in our example, with an important difference: although the card is removed from the catalog, the book (i.e., document) is not removed from the shelf. Instead, the map of shelf space is updated to indicate that the spot on the shelf is available. But the book (document) is allowed to sit on the shelf (i.e., in the computer’s memory) until another book (document) is added to the library and the spot is needed. With the large amounts of space on today’s computers it could be a long time before that spot is needed. Also, the “card” used to index the “book” in our example is not fully destroyed, but just marked as “gone.” The ability to identify documents that are marked for deletion (by reviewing descriptive document names) may make it possible for a computer expert to reconstruct other information about the document (such as the date it was created and the last time it was modified). This kind of information could be enough to tell an expert where to begin looking on backup tapes and other places for copies of the documents.

- Most organizations perform periodic backups of their data. Many organizations practice cyclical backups where data may be backed up every night, but to different media. The

media are often cycled in some pattern, such as on a weekly, monthly, quarterly and annual cycle. Sensitive documents may also be stored on these same backups. Thus, deleting a document from a computer server may really do nothing, as the document could be on dozens of backup tapes, especially if the document has existed for a substantial period.

- The problem grows worse because of the often rapid and uncontrolled dissemination of electronic data. Even if one user is diligent in deleting copies of a document, everyone on a “cc” list must do likewise or the document will not be safely deleted.

The implications are clear: even if computer users fully comply with corporate document retention policies, it is almost always possible to retrieve some of the data that has been “deleted” with conventional methods (more or less, depending on the resources and time dedicated to the effort).

There are several commercially available tools that can help with the “shredding” of electronic documents. “Disk shredders” typically attempt to address the persistence of supposedly deleted data by “overwriting.” Thus, swap file residue, deleted files and file names (any of which may contain all or part of deleted documents) are overwritten with random data. These tools are compared at the web site:

**[http://www.fortunecity.com/skyscraper/true/882/ Comparison\\_Shredders.htm](http://www.fortunecity.com/skyscraper/true/882/Comparison_Shredders.htm).**

Unfortunately, use of disk shredding software bears various costs and offers no absolute guarantee of permanent deletion. With the large disk drives in use today, the disk shredding process can be exceptionally time consuming. The more overwrites, the longer the process. Depending on the shredding software used, it could take anywhere from 30 minutes to 10 hours to shred data on a typical hard drive. During that time, the computer cannot be used.

Even overwriting data may not be enough. Experts have shown that magnetic traces may be recoverable (like “whiting out” a typewritten page and holding it up to the light to see the faint images still on the paper underneath the dried goo). It may be effectively impossible to sanitize storage locations by simply overwriting them, no matter how many overwrite passes are made or what data patterns are written.

For the total elimination of data, most computer professionals would advise following the practices of the United States Military. When classified information stored on any magnetic media must be disposed of, the physical media are destroyed (usually by melting down the data-carrying media).

#### Legal Development And Issues

Lawyers recognize that it is impossible to stop the accumulation of electronic records. Indeed, to a large degree, lawyers have embraced computer technology in their own operations. Electronic word processing, record storage and data sharing with clients all are generally seen by lawyers as a means to improve productivity and efficiency. Lawyers have also adopted electronic record search techniques as a major part of the way that legal research is performed. Law schools teach every law student the fundamentals of such research, and virtually every lawyer’s office has access to, and relies upon, electronic research capabilities. Many lawyers, moreover, have begun using sophisticated Internet and private networks as a means to harness and extend their data processing and analysis capabilities.

Lawyers have also extended the reach of electronic commerce. Lawyers, for example, have lobbied for the passage of electronic signature laws at the federal and state levels. These laws aim at making it possible to do business using purely electronic exchanges of contracts and other transaction documents. Lawyers have also obtained ethics opinions in many jurisdictions, holding that electronic exchanges between attorney and client can preserve privilege (thus

fostering such discussions). Far from holding back the creation of electronic records or insisting that only paper records have legitimacy, lawyers are generally in step with the modern electronic records based economy.

Lawyers, moreover, have long been aware that computer technology could have a profound impact on their dispute resolution practices. Rules of civil procedure, for example, have recognized since the 1970s that electronic records are “documents” that may be produced in litigation. Despite those rules, however, until recently, discovery in most litigation focused mostly on paper records. The reasons for this retarded development of electronic discovery are several.

First, even though many disputes involve businesses and institutions with extensive electronic records, many disputes really turn on a relatively small number of documents. A commercial dispute of modest size, for example, may involve a contract, some background information leading up to the formation of the contract, and some correspondence and internal memoranda regarding the performance of the contract. For disputes of this size, perhaps one or dozen documents are really key to understanding and resolving the dispute. These documents are generally easy to identify and retrieve (indeed, paper copies of the documents will often be stored in conventional records). Thus, sophisticated, detailed research into electronic records never becomes an issue.

Second, even though lawyers (and their clients) may be aware of the possibility that electronic records in their adversary’s hands may be important to a dispute, there is often a deep foreboding about the possibility of having to search one’s own electronic records for responsive materials. A kind of “mutually assured destruction” mentality develops, in which both sides of a dispute are deterred from pressing for review of all electronic evidence.

Finally, discovery of electronic records often involves substantial work and cost. In the first instance, if a requesting party wishes to obtain specific electronic records, standard discovery forms may need to be modified to reflect the request. If an adversary fails to produce electronic records, the requesting party will have to follow up, demanding information about what records exist and what searches were performed. If an adversary still resists production (or indicates that records no longer exist), the requesting party may need to get a court order to compel production of records, or to compel the adversary to give the requesting party's expert access to electronic records, for purposes of reconstructing and retrieving necessary information. And, when records are finally made available, they must be reviewed, analyzed and collated in a coherent form. At each step, cost, burden and time are involved. Unless the potential value of the records is high, the effort may not be justified.

These prevailing attitudes toward electronic discovery, however, are likely to change as lawyers and their clients come to realize that electronic records can dramatically affect the outcome of a case. There have been several recent, headline-grabbing lawsuits where electronic records (especially, e-mails) have served as vital evidence for one side. Awareness of the power of electronic discovery is likely to grow for several reasons:

- Electronic discovery offers the power to impose disproportionate burdens on parties in certain forms of litigation. Individual shareholders or consumers suing a corporation, for example, are likely to have no electronic records of their own, but they (and their lawyers) may be able to gain access to thousands (perhaps millions) of records in the corporation's files, with relatively little effort. The burden on the corporation of having to respond to such requests may be a powerful incentive for corporations to settle such suits, even if they have dubious merit.

- The risk of being charged with “spoliation” of evidence is increased in an electronic records environment. Spoliation, in most jurisdictions, generally means destruction of records where a party knows that the records may be important in litigation. Where a court concludes that a party has spoliated evidence, various sanctions may be imposed, ranging from imposition of the costs required to reconstruct needed information, to an “adverse inference” (basically, a ruling as a matter of law that the missing records may be presumed to have been supportive of the requesting party’s position), to an outright judgment in favor of the requesting party. In some instances, even criminal sanctions may accompany the destruction of records. Again, especially for a party with relatively few records of its own, demands for electronic documents may hold the prospect of either uncovering beneficial information, or at least making life very uncomfortable for the other side.

- Vendors are actively marketing technology to deal with electronic records, especially to lawyers. As with the recent Y2K crisis, a small army of experts and consultants is forming to exploit this market. Seminars on the subject are given throughout the country on a regular basis. Unlike Y2K, however, which resolved itself with a whimper rather than a bang, this “crisis du jour” is more than just a short-term economic opportunity for some get-rich-quick consultants. These consultants and experts are educating lawyers for new rounds of escalating conflict over electronic discovery, and there is no natural stopping point for such conflict.

- Increasingly, judges, even those raised before the dawn of the modern computer era, are becoming comfortable with the technology and the size of electronic discovery. Where once it might be sufficient to tell a judge that a request involved potentially millions of pages of records, a lawyer cannot assume that a judge will deny a request for discovery on volume alone.

More and more, judges are becoming aware that it is, in fact, possible to review masses of records at costs and on time frames that in the past would simply have been prohibitive.

The law in this area is still developing. Although the basic practice of discovery is familiar to most lawyers, practical solutions to many issues that can arise in electronic document production have yet to become standardized into any neat set of rules to govern most cases. Instead, rules for electronic discovery are largely developed on a case-by-case basis with each judge attempting to reconcile the needs and interests of the parties as they appear in the individual matter.

What seems certain, however, is that the issue of electronic discovery will not go away. Just as the trend in the data processing industry is toward ever greater volumes of information, the trend in law is toward ever greater demands for production of such information.

#### Implications for Business And Institution Managers

For many business and institution managers (and their lawyers) there is a tendency to fantasize about a magic solution to the electronic records problem. Surely, with our powerful technology, there should be some technical solution to the problem. A business or institution should be able to keep only its “good” records and discard, permanently, records that are harmful or useless. Yet, there is no magic solution to the problem. Most businesses and institutions keep vast and ever-increasing quantities of outdated and useless records, which, if anyone looked closely, include many records that could, in the light of litigation, be viewed as inappropriate, embarrassing or, in some instances, absolutely devastating.

If there is no magical solution to the problem, then are there at least some ways in which the problem can be contained? We suggest here a few basic principles to consider:

- Make records management a priority. As we have seen, the task of containing records is not an easy one. If no constraints are imposed, records tend to proliferate and to hide in many places. Substantial, active resources must be dedicated, if any real controls are to exist.
- Clearly identify categories of records that should not be retained. This task may be especially difficult, because of the numerous business incentives to retain records, outlined above. Review of legal constraints on document destruction, moreover, should be an essential part of this task. Failure to identify the target group of records, specifically, however, may result in over-inclusion or under-inclusion in the document disposal process. Employees (both those in information technology and those working in other areas) require clear guidance.
- Take steps to ensure that documents that should be destroyed truly are. In addition to considering technical solutions (electronic shredders and the like) the business may require specific policies and compliance mechanisms to ensure that copies of documents are not maintained in uncontrolled, unaccountable places within the organization. It may be particularly important, in this regard, to identify the “official” location for certain types of records, and to encourage employees to discard all copies of the records, other than those in the official location.
- Pay particular attention to concerns about privileged or highly confidential records. Although many of these records will have to be retained, clear identification of such records may make it much easier, if and when document requests are made, to ensure that privileged and confidential records are not produced, without some special consideration.
- Consider document organization and retrieval as a species of document retention. To a large degree, in today’s information environment, the contention that “we do not have any such records” is really a statement that “after reasonable search, we have been unable to retrieve any such records.” When a formal document request comes (in litigation, or as part of some

regulatory effort) the business will need to be in a position to establish that the search, in fact, was reasonable. If electronic records are well organized, and search capabilities are adequate, it will be much easier to explain (to a court or government agency) why the business should not be required to spend additional time and effort searching for records that might have been missed.

- Recognize that document destruction practices, once initiated, may be difficult to stop. Routine recycling of back-up media, for example, is an engrained part of the normal information technology function. There must be clear, effective contingency plans that will permit the business to preserve data whenever a dispute arises that may require use of the records, and certainly when a specific demand has been made for records, in a lawsuit, or by a regulatory agency.