e-Normous: The Increasing Burden Associated with Electronic Document Production in Second Request Investigations

Scott Sher and Daryl Teshima

The Problem

By now, everyone recognizes the increasing costs associated with the production of electronic discovery in the Second Request process.[1] The volume of electronic documents, including e-mails, PowerPoint presentations, memos and notes, is overwhelming, and increasing at a rate that puts Moore's Law[2] to shame.

At the same time, the Second Request process has remained largely the same, and the document demands from the Federal Trade Commission and the Department of Justice do not take into account the substantial burdens associated with electronic document review. The agencies still require parties to produce all potentially responsive documents, including the tremendous amount of accumulated electronic information maintained by companies in the ordinary course of business. As a result, parties are subjected to an increasingly expensive Second Request process they view as unnecessarily long and which prevents them from closing quickly and capturing efficiencies from their deals. On the other hand, the agencies require a process that ensures that they gain access to all important information from all relevant employees so that they properly can assess the likely competitive effects of transactions they review.

Although some in the Bar reflexively fault the agencies for the substantial burden of Second Request review, the responsibility lies both with the agencies and the Bar to streamline the process. After all, the agencies are not responsible for the massive accumulation of electronic data and cannot be expected to understand the complexities of every company's document retention and storage program. Only the parties themselves can educate the government as to where potentially responsive information resides and how the agencies can best gain quick and reasonably comprehensive access to it.

As a result, both the merging parties subject to Second Request investigations and the agencies must devise solutions that reduce the cost, time, and burden of producing electronic docu-

■

Scott Sher *is a partner at Wilson Sonsini Goodrich & Rosati, in the firm's Reston, Virginia, office, where he focuses predominately on antitrust issues of importance to high-tech companies.*

Daryl Teshima *is Vice President of Strategic Discovery, Inc., where he counsels law firms and companies on how to effectively utilize technology when responding to document and data discovery requests in civil litigation and government investigations.*

[1] Under the Hart-Scott-Rodino Antitrust Improvements Act of 1976, as amended, parties to any merger or acquisition of assets or voting securities exceeding a certain value ($53.1 million at the time this article went to press) must file a notification of the transaction with the Department of Justice, Antitrust Division, and the Federal Trade Commission. *See* 15 U.S.C. § 18a. By statute, the agencies have 30 days to determine whether to "clear" the deal and take no enforcement action, or issue a Request for Additional Information (commonly referred to as a Second Request) seeking additional documentary and narrative evidence regarding the competitive implications of the transaction. *See id.* The agencies use the Second Request process to determine whether any particular transaction will serve to "substantially lessen competition" in contravention to Section 7 of the Clayton Act. *See* 15 U.S.C. § 18.

[2] Gordon E. Moore's empirical observation that at current rates of technological development, the complexity of an integrated circuit will double every 18–24 months. *See* http://en.wikipedia.org/wiki/Moore's_law.

ments without compromising the integrity of the Second Request process.[3] In this article, we offer some observations based upon our experience and the data that we have collected over time. Most significantly, it is important for parties to share with the agencies, at the outset of any investigation, detailed information regarding the amount of electronic data, its nature, and its location throughout the company. Only with such a detailed understanding will the agencies be able to determine what information they need, and from whom. Once the agencies have such information, we believe that they should be in a better position to demand only what is reasonably likely to yield responsive and relevant information.

Analysis of Prior Second Request Electronic Document Productions

Because the cost of electronic document storage has declined so dramatically in recent years, companies tend to allow electronic data to reside on their underutilized storage systems for a very long time [and] there is the equivalent of tens of thousands of boxes of documents located in a company's electronic repositories.

In our effort to better understand the extent of the problem created by the substantial and growing volume of electronic documents maintained by companies, and to help craft some practical suggestions for how to resolve it, we compiled information relating to Second Request productions from the last five years. We compiled the data by collecting information from a number of Second Request investigations that we have managed for clients over time. We set forth some of our results regarding electronic productions from a representative subset of those deals. Because we primarily manage government investigations for high-tech companies, we acknowledge that it is possible that this information is skewed because, as some have suggested, high-tech companies make greater use of electronic document communication and tend to store more of that information. Anecdotally, however, many so-called "old-economy" companies make equal use of electronic documents and may have a *greater* volume of electronic documents due to retention programs that mandate maintaining employee data for an even greater length of time.

Whatever the bias introduced by examining only high-tech investigations, one thing is clear: because the cost of electronic document storage has declined so dramatically in recent years, companies tend to allow electronic data to reside on their underutilized storage systems for a very long time. Our clients maintain vast numbers of servers and storage devices that house nothing but old data—data that they no longer use but which is nonetheless potentially responsive to a Second Request investigation. Moreover, with the increasing automation of company-wide data—in the form of enterprise relationship (ERP) databases—our clients generally maintain a tremendous amount of information regarding their sales process in centrally located servers. As a result, there is the equivalent of tens of thousands of boxes of documents located in a company's electronic repositories. It is costly and burdensome for parties to review and produce all of this information, and the government is no doubt also burdened by reviewing such a large volume of material.

Some of the results from our study are highlighted below:

- The overall volume of electronic documents has steadily increased over time—from 0.73 gigabytes collected per employee in 2001 to 7.08 gigabytes collected per employee in 2005. This represents an increase of almost 970 percent, in just over four years. (See Chart 1.)
- Technology has improved over the last four years, greatly increasing our ability to quickly discard a subset of non-responsive documents, without a significant investment of time

[3] For a good discussion of the issues relating to FTC policy on electronic productions, see D. Bruce Hoffman, *The Digital Age at the FTC: Current Issues in Electronic Document Production and Review*, Antitrust Source, Mar. 2004, http://www.abanet.org/antitrust/source/03-04/hoffman.pdf.

and money. Nevertheless, even with aggressive document "culling," the volume of electronic documents that demand more careful review has increased steadily over time.

- Some categories of information—deleted e-mails, for example—demand an incredible amount of review time without yielding much in the way of responsive information. There has been no instance in our experience where a deleted item was significant in an investigation. In fact, even though deleted e-mails constitute over 21 percent of e-mails that need to be reviewed, they only amount to 0.45 percent of "responsive" information.
- While the volume of information collected from sales employees is huge, such information is far less likely to be responsive than information maintained by company executives.

We analyzed information related to electronic document collection, review, and production from five deals that we managed during the period beginning in early 2001 and ending in early 2005. Each of the deals discussed were subject to a Second Request—and reviewed by either the FTC or DOJ.[4] Each involved roughly the same Second Request demand: asking for responsive data from the companies for a period of approximately three years. We provide information related to the collection, review, and production of *all* electronic documents pertaining to these investigations, not merely e-mails and not only stand-alone electronic documents. We recognize that the results of our analysis can only be used to inform our discussion in an anecdotal, not scientific, way.

Methodologies for Collecting, Reviewing, and Producing Electronic Documents

Collecting Electronic Documents. Perhaps the most difficult and overlooked aspect of electronic discovery is the burden of collecting information required by the Second Request. Because the agencies rely on the Second Request as their primary opportunity to obtain documents during a merger investigation, the scope of the Second Request is intentionally broad and often applies to all business-related documents and data within the organization.

The casting of such a wide discovery net historically has been tempered by the limitations inherent in storing paper. There is a limit to how much square footage a company can devote to housing paper documents. Electronic data and documents do not share these same physical and cost barriers. In addition, the requirements imposed on companies by Sarbanes-Oxley[5] and other federal and state regulations[6] have fostered a "keep everything" environment. All of these factors

---

[4] The five deals surveyed by this article are:

- Deal "A"; October 2001 (Predominately a Hardware Deal; Reviewed by FTC Mergers II)
- Deal "B"; March 2003 (Software / Services Deal; Reviewed by DOJ NetTech)
- Deal "C"; June 2003; (Chip Deal; Reviewed by FTC Mergers II)
- Deal "D"; July 2003; (Software Deal; Reviewed by SF DOJ)
- Deal "E"; January 2005; (Software Deal; Reviewed by FTC Mergers II)

[5] Pub. L. No. 107-204 § 802(a).

[6] SEC regulations are the primary source of statutory requirements that mandate document retention. For example, section 802(a)(1) of the Sarbanes-Oxley Act requires auditors of public companies to maintain "all audit or review workpapers" for five years, and directs the SEC to enact related regulations. Thereafter, the SEC amended Reg S-X by adopting Rule 2-06 that established a seven-year retention period for "records relevant to the audit or review, including workpapers and other documents that form the basis of the audit or review, and memoranda, correspondence, communications, other documents, and records (including electronic records), which (1) are created, sent or received in connection with the audit or review, and (2) contain conclusions, opinions, analyses, or financial date related to the audit or review." 17 C.F.R. § 210.2–06(a). In addition, many companies are under document retention orders as defendants in private litigations; this too adds to the volume of electronic documents stored by firms.

have resulted in a perfect storm of overwhelming electronic data volumes that, in turn, are often demanded as part of the Second Request process. Over the past four years, we have seen this increase firsthand. In 2001, we collected from our clients the equivalent of 14.6 boxes per employee of electronically sourced documents. As of 2005, this number had increased almost tenfold, with an average of 141.6 boxes collected per employee.[7] (See Chart 1.)

Clients increasingly are unwilling to absorb a corresponding tenfold increase in Second Request compliance costs and fees, and often look to their legal counsel for suggestions on how to control the costs associated with this expanding document universe. One way to control the volume increase is through a more discriminating collection methodology. Traditionally, three strategies for collecting electronic data during Second Requests are employed:

1. Assign collection duties to the company's internal staff, such as the information systems group;
2. Hire forensic collection specialists who make images of hard drives and servers that contain every bit and byte from a employee; or
3. Assign collection duties to individual employees to self-select responsive electronic documents.

Over the past four years, we have seen this increase firsthand. In 2001, we collected from our clients the equivalent of 14.6 boxes per employee of electronically sourced documents. As of 2005, this number had increased almost tenfold, with an average of 141.6 boxes collected per employee.

Unfortunately, these traditional strategies often result in the collection of either too much data (the continuing problem of the over-collection of non-responsive documents) or too little. Both the use of forensic imaging of hard drives and the sole reliance on a company's internal IT staff to collect electronic documents usually result in over-collection. Forensic imaging copies not only user-generated data, but also system and program files that do not require review. Similarly, a company's IT staff uses tools that are geared toward data recovery, which results in the collection of duplicative and non-user-generated electronic files. Conversely, relying primarily on the individual employee to identify responsive data often results in the collection of too little data (especially considering the broad scope of Second Requests). As a practical matter, it is also difficult to get busy employees and executives to take the time to properly identify data that is potentially responsive to the Second Request.

Despite these disadvantages, there are benefits to each of the above collection methodologies. A company's IT staff has unique access and knowledge of internal systems, allowing them to develop the most efficient and least intrusive method of collection. Alternatively, utilizing forensic tools and procedures properly preserves metadata and the chain of custody for the file. Last, no one better understands the substance and significance of the data to be collected than the custodian of that data. A hybrid approach accordingly incorporates the best elements of each strategy. Generally, such an approach follows these steps:

- **Catalog Data Universe**—The first step is to determine the various possible data sources at the company. This is usually done in conjunction with the company's IT staff. Generally, four main data stores are useful targets:
  - *Individual Employee Files*—These files are controlled by individual employees on their computers, e-mail servers, or private network shares. The scope of these files could extend to an employee's home computer, personal data assistants, and other devices, depending on how an individual employee operates.

[7] These metrics reflect the total amount of user-generated data collected for each employee. It does not include system and program files, which were excluded from the collection specification.

- *Department/Group Files*—These files are controlled by a department or select group at the company and are typically stored on the company's network server.
- *Enterprise Databases*—These company-wide systems contain personnel records, accounting and financial information, customer and sales records, and other business information. Due to the complexity of these systems, they are usually maintained by the company's IT staff and access is limited to a select group. Most users access these systems through regularly-generated reports.
- *Backup Tapes*—Usually stored on removable media, this data is comprised of information that the company regularly backs up and is controlled by the IT staff. Depending on the electronic data retention policy (and tape backup recycle system) in place, hundreds of these backup snapshots are conceivably available. In general, this data is often highly duplicative of data found in the other three data stores.

- **Interview Employees (and learn how they organize their e-mail/electronic documents)**—A collection interview should be set up with individual employees in order to: (1) allow the counsel to evaluate each employee; (2) collect electronic data controlled by every employee (e.g., such as data on laptop, computer, etc.); and (3) verify that all relevant data from each employee is being collected.
- **Sweep Data**—Counsel should arrange to "sweep" an employee's hard-drive during the employee interview (or remotely, over the network) to find material that is likely to be relevant, based on objective criteria, such as file-type, date of file, and so on. Files that meet the criteria can be copied by a program onto the collection media. The sweep collection can also include a topical culling stage where the employee identifies data that will be extremely unlikely to be responsive (such as personal folders). The sweep method, if performed properly, can also properly preserve all metadata and make the collection more forensically sound.
- **Compile Metrics**—After the collection is completed, counsel should determine the volume of data collected for each company employee. To do this, one must accurately translate the amount of data collected into pages, which will give the attorneys, clients, and agencies a measuring unit that they can readily understand. Using these metrics, descriptive charts can help clients and the agencies understand the nature and composition of the collected data and, ultimately, identify the documents most relevant to the Second Request.
- **Prepare Descriptions and Reports from Key Enterprise Databases**—Unlike individually maintained electronic files and e-mails, simply copying (or printing out) databases will not likely generate information that can be meaningfully reviewed. Likewise, the detailed delivery of database schemas and descriptions that we have seen in several Second Requests does little to generate usable information. The best approach is to create a high-level description of the database, as well as a list of the relevant reports generated and distributed by these databases and their applications. By focusing on producing or generating reports (or data exports), counsel can then use these reports as the basis for negotiations with staff and use them as a method to modify or satisfy the burdens of the Second Request.
- **Verify that Backups Contain Duplicative Data**—The large volume of data collected calls into question the relevance of collecting information from backup tapes or other disaster recovery media. Backup tapes are often highly duplicative of e-mails and electronic files already collected and are quite expensive to restore. Further, because backup tapes are

made for disaster recovery purposes by IT personnel and are designed to restore all data at an enterprise, rather than the individual files of a particular person, the recovery process is complex.[8] In order to avoid this process, parties should verify (usually by the date range of files and e-mails collected) with the agencies that all responsive information has been collected without the need to restore backup tapes. Absent a date gap in the collection or a situation where the backup contains the only copy of material (if for example, an employee's hard drive was erased completely and never restored), restoration of backup tapes may be unnecessary and avoidable.

Reviewing Electronic Documents. Unfortunately, even following a hybrid collection approach[9] does not guarantee that the parties will have a manageable universe of potentially responsive documents to review. As previously mentioned, an average collection can easily yield the equivalent of over 140 boxes per employee to review. It is not possible to review this volume efficiently or in a timely manner without excessive attorney-review costs. Thus, counsel should take additional steps to further narrow the universe of potentially responsive documents

The logistics of completing a hard-copy review are ill-matched to the mountains of data generated by electronic discovery.

Traditionally, documents are reviewed in hard-copy paper form, which is often the most familiar and comfortable method for attorneys. But the logistics of completing a hard-copy review are ill-matched to the mountains of data generated by electronic discovery. First, printing electronic documents is an expensive proposition, and can cost anywhere from $0.06 to $0.12 per page to print, depending on the type of data to be printed—and that is for just one copy.[10] Second, the only way to review paper is page-by-page. Since the typical digital discovery collection can average millions of pages, resource (or budget) limitations may not permit adoption of a paper-based approach.

In response to issues raised by a paper review, litigation review databases have proven to be a common, cost-effective solution to reviewing electronic discovery. There are hundreds of service providers (and even off-the-shelf programs) that can take electronic files and e-mails (at a cost ranging from $0.08 to $0.25 per page) and perform two basic functions: (1) extract all the metadata and full-text into a searchable database; and (2) generate corresponding electronic printouts (usually in .tiff (or .tif) or Adobe .pdf formats) of the electronic data that can be linked to the extracted searchable text. Once processed, the searchable text and images can be loaded into a database as well as made available through the Internet by a number of application service providers (ASPs). All of the above litigation support programs help attorneys reduce the cost of photocopying yet provide the option to print paper sets at any time. In addition, everything that can be done during a paper review (e.g., capture attorney calls, redact documents, organize into groups) can likewise occur electronically.

Taking electronic files and e-mails and placing them in a searchable database format can reduce by as much as two-thirds the cost of transforming paper into a similar searchable data-

[8] To recover a particular user's files and e-mails from backups, the entire set of data must be restored. Only after the data is restored can individual e-mails and electronic files be reviewed.

[9] This hybrid collection approach represents a smaller subset of the entire data universe controlled by an employee, and does not include non-user generated files (such as Windows system and application files) or files identified by the employee as non-responsive (such as folders where personal e-mails and documents are kept). By using these filters at the time of collection, the total volume of the collection is usually reduced by 50–70 percent.

[10] Paper-based reviews usually require at least one pristine copy, multiple working copies, and a final production set that is properly Bates-numbered.

base. Attorneys further minimize the number of paper copy sets generated, often resulting in considerable cost savings.

And over time, the searchability of a database helps attorneys achieve faster paper-review speeds and allows the parties and ultimately the agencies to identify and filter out clearly non-responsive documents. If searchable, electronic data can often be effectively narrowed utilizing the following culling criteria:[11]

- *Date Range*—If there is a relevant date range for the matter, then e-mails and files can be filtered to ensure that they fall within the applicable time period.
- *Duplicates*—Using a variety of methods, duplicates can be objectively identified and ultimately removed from the electronic data collection. An important factor is the universe of documents from which documents are eliminated
- *File Types*—Electronic data that is not user-generated (e.g., system files, applications) or yields poor results when printed / converted (e.g., databases, multimedia files) can be logged and not processed.
- *Keywords*—Search terms and names can effectively separate the wheat from the chaff. In practice, however, we have found that negotiating keyword search terms with the agencies is a time-consuming process because it is difficult to reach agreement on which terms to utilize.[12] Such a negotiation process often requires a significant amount of time and explanation, and eventually results in a slight reduction in the overall volume to be searched. We do not blame the agencies for this process—after all, from their perspective, it is impossible to know what is lost from review if too narrow a universe of search terms is agreed upon. None of the Second Requests surveyed in this article utilized keyword searches to eliminate documents to review.

For each of the Second Request productions we surveyed the data was put into a litigation database and reviewed online over the Web. This gave the counsel's review team (situated in different locations) a centralized system by which they could search all documents as they are being reviewed, as well as the ability to cull documents using some of the criteria discussed above. As technology has improved, the ability to cull has likewise dramatically increased. For example, as discussed below, we were able to reduce total volume by over 70 percent in our most recent Second Request. But even improvement in filtering and culling techniques has only allowed us to tread water, as the amount of documents we have had to review after filtering has increased 300 percent since 2001.

---

[11] On December 11, 2002, the FTC published its Statement of the Federal Trade Commission's Bureau of Competition on Guidelines for Merger Investigations, which discusses best practices with regard to the production of electronic data. *See* Federal Trade Comm'n, Statement of the Federal Trade Commission's Bureau of Competition on Guidelines for Merger Investigation §§ 6.b and 6.d (Dec. 2002) [hereinafter FTC BC's Merger Investigation Guidelines], *available at* http://www.ftc.gov/os/2002/12/bcguidelines021211.htm. Among other things, these Guidelines acknowledge the efficiencies gained with electronic productions and contemplate the use of term searches to cut down the volume of e-mails.

[12] FTC BC's Merger Investigation Guidelines contemplate the use of term searches to help reduce the volume of documents that need to be reviewed by the parties. *See* FTC BC's Merger Investigation Guidelines §§ 6.b and 6.d, *available at* http://www.ftc.gov/os/2002/12/bcguidelines021211.htm. These guidelines also acknowledge the difficulty in negotiating term searches because it is difficult to negotiate searches without understanding the documents and the industry. In order to negotiate term searches, the Guidelines recommend that the responding party provide: (1) organizational charts and background on how electronic data is kept at the company; (2) glossary of industry and company terminology; (3) proposed search methodology and search samples and results for staff to evaluate; (4) rolling production so that search terms can be adjusted; (5) exceptions from searching for certain key employees; and (6) safeguards to ensure that responsive documents found outside searches will be produced, or not relied upon by the party. As one would suspect, such negotiations are cumbersome, and in our view, actually serve to hinder the pace of review.

Producing Electronic Documents. The only reliable constant in Second Request production formats is unpredictability.[13] For each of the surveyed Second Requests, the government has requested different production format combinations, which include:

- **Access to Online Review System**—Online review gives the agencies Web access (usually the same one used by us during our review) to the universe of documents that the parties produce in response to the Second Request. To address concerns about reliability, performance, and security of the online review system, we have agreed to enter into service-level agreements with the agency, whereby the agency maintains control over the data and the parties have no access to the agencies' online review.
- **Paper**—Paper is the traditional method of production, and requires the parties to print out each electronic document, which is impractical, expensive, and kills trees.
- **Database and .tiffs**—In some instances, the agencies demand production of documents in a database—usually Concordance or Summation—that contains full-text and non-work-product metadata linked to a .tiff electronic image of the document (similar to a black and white photocopy).
- **.pdfs**—The agencies have also asked for production of documents in searchable Adobe Portable Document Format (.pdf) (which is a color, searchable photocopy).
- **Native Files**—The agencies also sometimes accept a copy of the original files from the parties without converting the files into an electronic image or paper. To date, we are not aware of any instance where the agencies have accepted a production comprised entirely of native files, but instead, require native file formats only where the agencies seek to have the ability to "modify" the document and manipulate the core assumptions in the document (e.g., Excel spreadsheets).

Although each of the production formats has various benefits and disadvantages, all are sufficient for giving the agency an opportunity to review and assess the production. While we generally are production format agnostic, the most significant non-attorney expense is the cost of processing the collected data into a common-denominator format, whether paper or an electronic image. This conversion process also takes a significant amount of time, significantly delaying the review and production process. Accordingly, there is increasing pressure from clients to skip this conversion step and instead review and produce just using the native files (e.g., documents that have not been converted into searchable reproductions such as .tiff or .pdf images). A native-file review was very difficult to execute a few years ago due to the limits of available technology and tools, but significant developments have made this process easier.

Because electronic-file volumes likely will not decrease in the future, the ability of the agencies to accept a native-file production (especially because these are how the files are kept in the ordinary course of business) will become critical. At present, however, it is difficult to negotiate a native-file production with the agencies. We leave for another day the debate as to whether the HSR Act allows the agencies to demand a specific format for production or whether production of documents in native-file format would somehow not constitute "substantial compliance" under the Act.[14]

---

[13] In fairness to the agencies, there is no standard production format in the litigation support industry. As with any emerging technology industry, new applications and services emerge every day. To their credit, the agencies have tried to implement standards. *See* Hoffman, *supra* note 3.

[14] 15 U.S.C. § 18a(g)(2) (setting forth substantial compliance timeframes). There has not been any formal or informal guidance, to date, as to whether native production of electronic documents would constitute substantial compliance.

The problem created by the uncertainty surrounding production formats is that it often limits choices in the selection of litigation support vendors that will generate the production sets. That vendor needs to have a scalable online review system, as well as the ability to turn around the large production volumes (in a variety of the above production formats) generated from a Second Request review. Decisions regarding production formats also tend to come late in the negotiations between staff and the responding party, which makes it hard for the responding party to plan efficiently. The lack of clear production guidelines also can result in deliveries that do not meet the technical specifications of the agency, further delaying the entire process. Whatever the format, the establishment of consistent, reasonable, and clear production guidelines will make the process more efficient.

Results of Our Survey

Observation #1: The volume of electronic documents is expanding exponentially.

In the deals that we examined, the average volume of documents collected per employee has increased at an almost unbelievable rate. We observed that from 2001 until early 2005, the average volume collected (converted from gigabytes to estimated boxes) per employee jumped from 14.6 boxes to 141 boxes, an increase of over 969 percent. (See Chart 1). To put that into the context of complying with a Second Request, the same number of attorneys required to review documents from 14 employees in 2005 would have been used to review 135 employees' documents in 2001. That alone has increased the cost of the Second Request process tremendously.

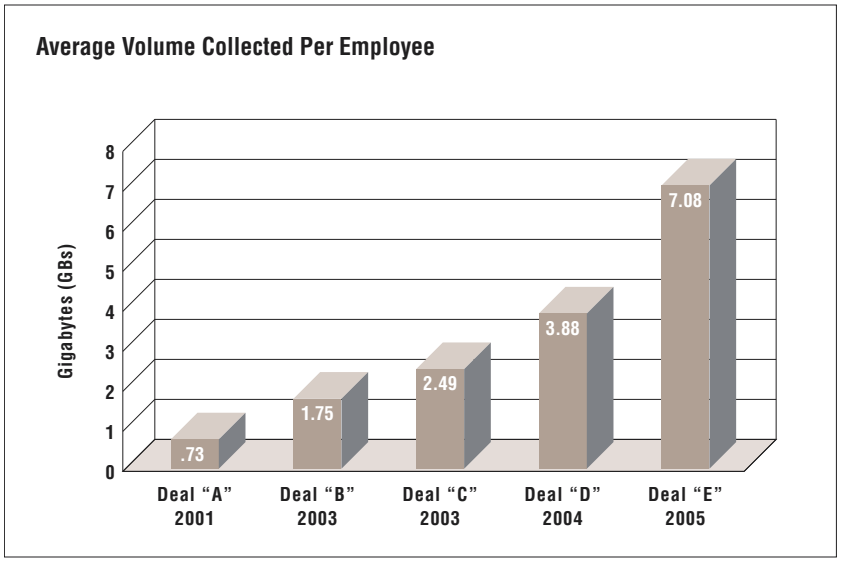

Chart 1: *For the five deals surveyed, this chart tracks the average volume of data collected per employee. In general, one gigabyte of data is equal to 50,000 pages (or 20 boxes).*

Observation #2: E-mail represents the single largest source of electronic documents (see Chart 2) and in all likelihood accounts for the most substantial component of electronic discovery cost in Second Requests.

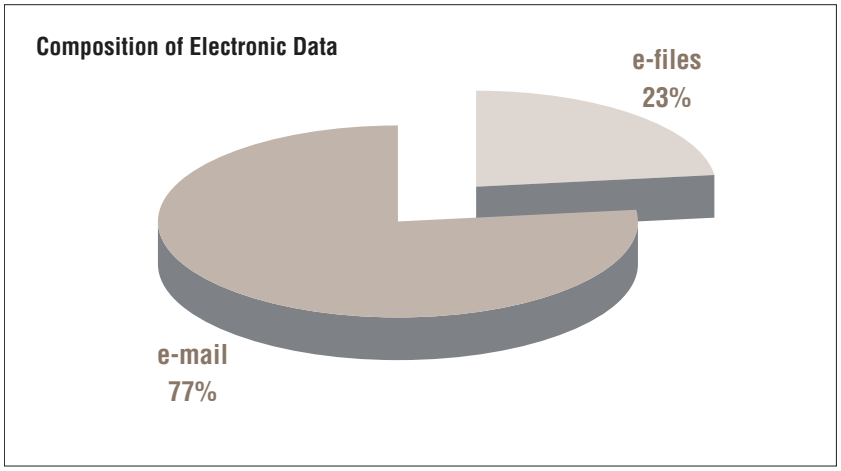

Chart 2: *For the five second requests surveyed, the vast majority of data collected was e-mail.*

For our clients, at least, e-mail has become the most important method by which they communicate–with their work colleagues, customers, suppliers, and lawyers. E-mail is an easy way to communicate information quickly and concisely, and for many of our clients, it simply never goes away. It is not uncommon these days to open the e-mail folders of the Vice President of Sales and find *several hundred thousand* e-mails in her inbox and the same number in her deleted items folder. These e-mails are unsorted—the equivalent of dumping several hundred or thousand boxes of documents on the floor. Multiply that by the number of employees that the parties must search to respond to a Second Request and it is clear how quickly the burdens can become enormous.

Observation #3: Technology improvements have made it easier to "cull" non-responsive documents from production (see Chart 3); however, the improvements do not keep up with the increasing volume of documents. As a result, culling does not solve the problem of burdensome document reviews.

"Culling" electronic documents is the process of discarding electronic documents that clearly are not responsive to a Second Request, without investing a significant amount of attorney time. Given the volume of electronic data, culling most importantly includes the automated removal of documents that fall outside of the responsive date range in a Second Request, or the elimination of duplicates.

We have also increasingly turned to other methods of "electronic" culling. For example, prior to a detailed review of documents, we often manually review the file trees (i.e., an electronic directory of each individual's electronic documents) collected from each employee from whom we need to collect in order to respond to the Second Request. Where file trees clearly contain non-responsive information (e.g., a file folder entitled "Johnny's Third Birthday" or a file folder containing information regarding products not included within the scope of the Second Request), we eliminate such documents from the review process. This culling has a large impact on the scope of our actual detailed review.

Even with aggressive culling procedures, we are still reviewing a significantly larger universe of documents. (See Chart 4). Note that in Deal A, in 2001, we had not yet developed a reliable method to cull information, so we manually reviewed every electronic document collected from each employee whom we searched at the company. By 2005, we were able to eliminate a full 70

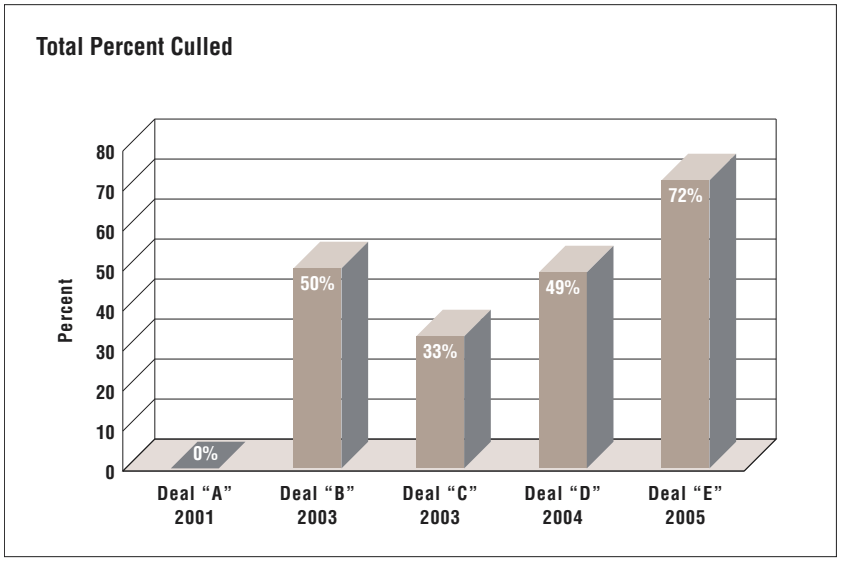

Chart 3: *This chart demonstrates the percentage of documents that we have been able to cull from the review universe utilizing a variety of technologies and techniques.*

percent of documents that we collected without conducting an extensive review of every document within each employee's electronic document cache. Nevertheless, even with aggressive culling methodologies, we are still reviewing a tremendous amount of information—more than three times more than we reviewed in 2001, before reliable culling methodologies were adopted.

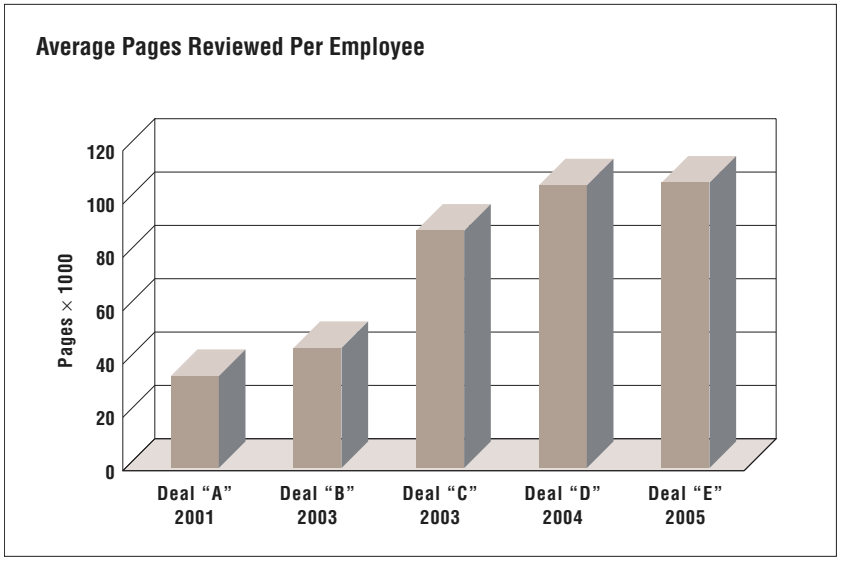

Chart 4: *By employing more aggressive culling techniques, the average number of pages reviewed per employee has plateaued for the time being.*

Chart 5 indicates that, even with aggressive attempts to eliminate clearly non-responsive information without detailed review, parties are still forced to expend a significant amount of time and money on review of clearly non-responsive documents. In a recent Second Request (Deal E), only one-third of electronic documents that were reviewed after eliminating clearly non-responsive documents were actually responsive to the government's Second Request demand.

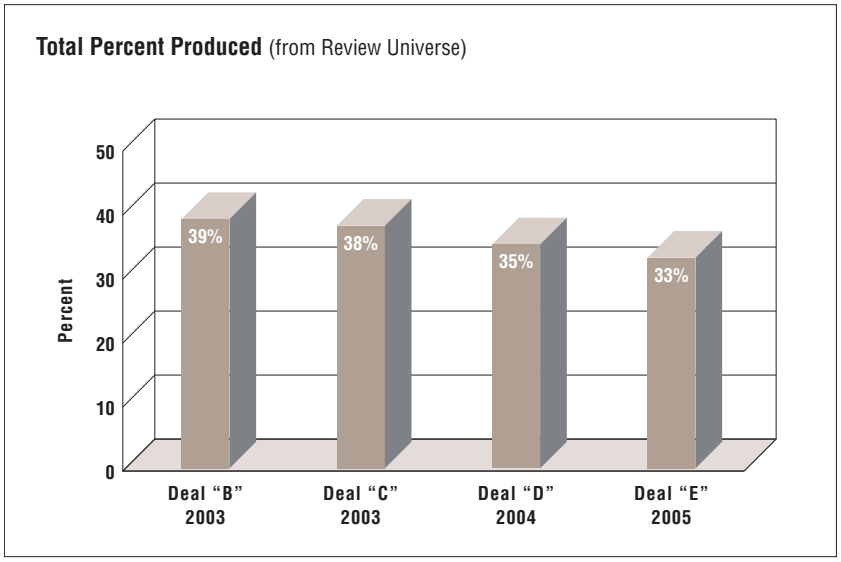

Chart 5: *Even with more accurate culling techniques, there is still a great deal of non-responsive data that needs to be identified by the review team.*

Observation #4: Management has the highest percentage of responsive documents; the sales force has a high volume of documents of questionable value. (See Chart 6).

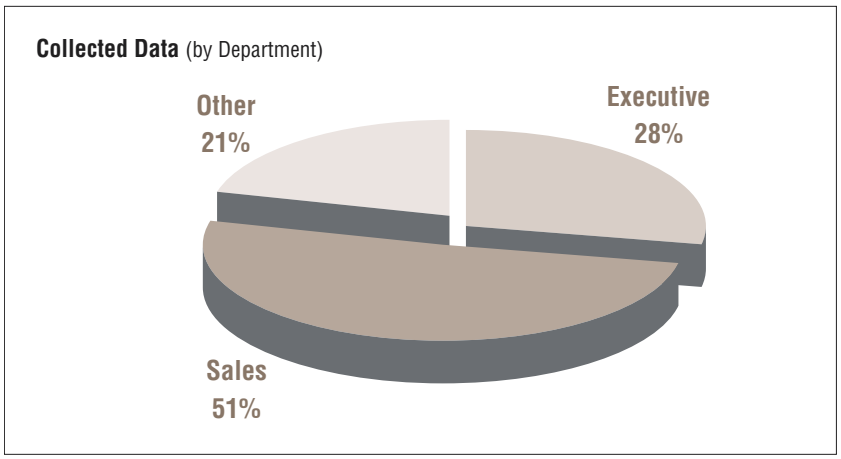

Chart 6: *The largest source of documents is from the Sales department.*

The sales force of a company tends to send and receive the most e-mail. There is often a considerable amount of chatter between sales representatives, some of which indeed may be "technically responsive" to a Second Request—for example, banter regarding a particular potential customer. We agree that such information may indeed be valuable. But especially where such customer-specific data is captured in centrally stored enterprise databases, we question whether the production of additional data adds anything to an agency's investigation. Further, older "banter" (e.g., from three years before the merger's occurrence), seems so distant as likely to be irrelevant to analyze the current state of competition in a market.

Recommendations

Recommendation #1: At the outset of an investigation, the parties should provide the agencies with detailed information regarding the content of electronic databases. In return, the agencies should agree to reasonable limitations on the production of electronic documents.

At the start of any investigation, the parties should make available to the agencies the individuals responsible or most knowledgeable for the maintenance of the companies' electronic documents. If necessary, the parties should agree to a deposition of such individuals at the outset of the investigation, or agree to answer "contention-type" interrogatories regarding the nature of their electronic document systems. As discussed above, the parties have available to them sophisticated tools and visualizations that quickly enable them to identify electronic document volumes and sources. These tools can be critically important in the Second Request negotiation process.

Recommendation #2: The agencies should be amenable to multiple production methodologies to limit the costs associated with Second Request production; the parties should agree to provide the agencies with the information necessary to complete an expeditious investigation.

As discussed above, there are numerous ways to produce documents—in paper format, natively, in Summation or Concordance, or online via a secure ASP. The agencies should work to overcome their reluctance to accept documents in multiple formats *and should accept any format*, so long as the parties are at the same time willing to certify that production via any methodology will not prejudice the agency's review. Often, we have negotiated with the reviewing agency concerning the production methodology and have promised "back-up" production for certain documents, if the methodology initially selected failed to produce all necessary information in a format that enabled the agency to complete an adequate review.

Recommendation #3: The agencies should pare down the number of sales employees searched and produced; the parties should promptly provide information regarding the type of documents maintained by each employee.

There have been calls to statutorily limit the number of employees from which the agencies can demand documents. We believe that such artificial limitations are not helpful and a one-size-fits-all approach is not appropriate. A merger involving two organizations, each with 30 employees, naturally demands a more limited production than a merger involving two organizations, each with 5,000 employees. Setting a fixed number of employees from which the agencies can search files helps no one.

Rather, the agencies must carefully consider the nature of their request. It goes without saying that the FTC and DOJ must focus their document demands. Perhaps, as some have suggested, if the agencies were required to publicly produce statistics to Congress regarding the metrics of document demands and production, and were required to meet certain performance objectives, this would create sufficient incentive for staff to temper their document requests.

The staff cannot reduce their demands, however, if parties require them to throw darts blindfolded. The parties must provide information early in the process regarding the nature of each employee's job responsibilities, and must take care to document carefully which employees likely will have the most responsive information. Anything less than that will contribute to the problem—and the parties cannot then blame the staff for burdensome production demands.

Recommendation #4: The agencies should limit the time scope of e-mails to be searched very carefully and should consider eliminating "deleted" items from Second Request productions; the parties should provide "representative samples" of such documents to the agencies.

Merger investigations are not conduct investigations. There likely will not be a smoking gun in someone's e-mail describing the effect of the merger on the market, and we have never seen the random e-mail from someone in field sales that ex ante described the competitive effects of consolidation on the market. In conduct investigations such smoking guns may exist—there may very well be the e-mail that describes a price-fixing scheme—but that is not the world of merger investigations.

As a result, we believe that the percentage of truly relevant e-mails is low. The percentage of ground-breaking e-mails is even lower. The percentage of deleted or unsorted in-box e-mail from two years prior to the contemplation is largely irrelevant. The burden to review such information is enormous. As the data shows, e-mails are the bane of merger review, and the process of collecting, reviewing, and producing such information is staggering, especially relative to its actual value. It is critical that something be done to control the e-mail production process. The agencies should consider seriously from which employees e-mails are relevant and the parties should quickly provide statistics to the agencies regarding the volume of such information.

Recommendation #5: The agencies should streamline the privilege log process.

Besides producing responsive documents, another significant production task is the generation of a privilege log. The amount of detail for the production log requires the parties to devote significant (and senior) review resources. While the parties can take advantage of document metadata (such as author, recipients, date, document type) associated with the document, and capture the necessary review basis and description during the online review process, the entire process is cumbersome and usually is the last item that is completed before certifying compliance with a Second Request.

Although we recognize the need to justify a decision to withhold a document on the basis of privilege considerations, we do not believe that the current process for privilege log production is rational. We believe that there is a middle ground that can streamline the process. For example, the agencies should simplify the level of detail for the production log and eliminate much of the data requested regarding each document in the production. The agencies could also limit the creation of a production log to a subset of documents (such as for only a few employees to be named by the agency). This would provide the agency with the ability to spot check the privilege calls made by the party and reduce the resources required to generate the privilege log. Another possibility—and one that we strongly believe is warranted—is the elimination of in-house counsel from the demands of a Second Request in all but the most limited of circumstances (e.g., where the General Counsel is also responsible for corporate development or has access to lawsuits that are relevant to the investigation, or has other necessary information that cannot be gleaned from non-attorney sources). It goes without saying that counsel is likely to have a tremendous amount of privileged data, and the cost and time associated with logging such information in a cumbersome log simply is not justifiable where the agencies can satisfy their need for information from other sources. ●