

The Size, Scope and Growth of the Electronic Data Discovery Market: Survey and Results

George Socha
Thomas Gelbmann

Introduction

In an industry that knows how to keep its secrets, one team is trying to find the answers to one of litigation's persistent questions – what is the size, scope, and growth of the electronic data discovery market?

In early 2003, we conducted a snapshot survey of the electronic data discovery (EDD) market, interviewing over two dozen vendor, law firm, and corporate legal department personnel and scouring on-line resources. We had three objectives:

1. Approximate the size of the EDD market as a whole, its growth over the past 10 years, and assess its potential for future growth;
2. Define and describe the major and minor segments of the EDD market and assess their past, present and future; and
3. Identify the leading providers and estimate the market shares of the top 5 EDD vendors.

Results of our survey indicate the 2002 domestic, commercial EDD market was at least one-quarter of a billion dollars, up from about \$40 million in 1999. That's a six-fold increase in three years.

Expectations are that the market will continue to grow exponentially for the next couple of years, with growth fueled by the changing ways in which information is created and stored rather than by any increase in litigation. We anticipate that growth will begin to slow by 2005-2006, tempered by more effectively electronic records management practices, improvements in technology, judicial intervention, and more targeted discovery demands.

From our vantage point, there are 3 main segments in the EDD market – “computer forensics/investigative,” “electronic discovery/litigation support,” and “consulting.” Vendors are flooding into the first two segments; we have over 100 in a list that we maintain and it keeps getting larger. We expect, however, to see an increasing amount of winnowing and consolidation as the market tests vendors' commitment to this arena. We anticipate that a core group of vendors will stay the course and emerge as dominant players. Others, however, will drop out or be absorbed by competitors, having failed to realize short term profitability, maintain focus and commitment to the market, or develop and pursue a viable long-term strategy.

Within the computer forensics segment, we saw 5 sub-segments developing: data identification, data collection, data restoration, data conversion, and data culling and

filtering. With the electronic discovery segment, we identified 4 sub-segments: data hosting, filtering and searching, data analysis, and search software. We did not observe that same focus on consulting. We viewed both consulting and efforts to develop standards as parts of the EDD market that are significantly underserved.

Among the 100+ EDD vendors in the market today, only a handful command quick recognition. Even among most visible, no one vendor emerges as the dominant player. In the next tier, there is a great deal of compression, with little effective differentiation between the vendors and the services they offer. Past the first 15 or so vendors, none appear to have much name recognition.

As you read on, keep in mind that the information and conclusions reported here are the results of a preliminary study. This is a first look into the size, scope and growth of the EDD market, not a definitive analysis. Our intention was to take a snapshot and draw an initial picture of the EDD market, trusting that the results would give us guidance on aspects of the market that deserve a closer look.

Methodology

Define the Scope of the EDD Market

Our first step was to define the scope of the EDD market. For purposes of the study, we confined our examination of the EDD market to the following areas:

1. “Born electronic” materials – We included activities connected to materials that were “born electronic,” those materials that started out in electronic form and stayed in electronic form until they entered the EDD pipeline. We excluded activities directed at gathering, processing, and working with materials that were not in electronic form until they entered the pipeline. We also excluded materials that were “reborn electronic,” ones that started in electronic form, got converted to paper, and then were re-converted to an electronic form by vendors.
2. Domestic activities – With our limited scope and considering that this was an initial snapshot, we confined our study to EDD activities within the United States.
3. Commercial activities – We looked only at EDD vendors who catered primarily to the commercial market. Consequently, EDD activities associated with governmental work and vendors who cater to that market are not in this study. We recognize, however, this area may be open to further study.
4. Activities from 1993 on – We used 1993 as a chronological starting point because prior to then there were not enough EDD vendors or activity to provide meaningful information.
5. The limits of the EDD market – We did not include the more traditional automated litigation support activities such as document scanning and coding.

Develop Preliminary EDD Market Segment Definitions

Having set the scope the EDD market, we then developed a set of preliminary major and minor EDD market segments. We initially described three major segments – software, electronic discovery services, and consulting services. We divided those principal areas into smaller ones. For example, the electronic discovery services segment included data identification and acquisition, data conversion, processing and pre-filtering, data delivery services, and data hosting. During the course of the study we continuously refined each segment.

Develop Survey Structure

As appropriate for this level of study, we conducted a telephone survey. We targeted three categories of interviewees. On the provider side, we sought interviews with EDD vendor representatives. On the consumer side, we contacted law firm and corporate legal department personnel.

We prepared three series of interview questions, one for each category of interviewees. While many of the questions were the same across all categories, others were designed to elicit information unique to a specific category. Many of the questions simply called for a “yes,” “no” or “I don’t know” answer. A large percentage of the questions, however, called for narrative responses. A few of our questions were open-ended; we phrased these queries so as to avoid influencing responses.

We entered the responses into a series of spreadsheet worksheets. By doing this, we were able to rate and weigh the results during the analysis and synthesis stages of the project.

Get Information from Secondary Sources

While we designed the survey questions, we also gathered information from secondary sources. These sources included vendor websites, published results of related surveys, white papers, listings of corporate legal departments, large law firms, and major litigation departments, and reports on judicial caseloads. By the end of the study, we had collected and evaluated information from over two dozen sources - nearly 100 MB of secondary source data.

In addition to gathering information about EDD vendors, we also collected data about the history of the EDD market, the total potential size of the EDD market now and for the next few years, and the total addressable EDD market.

Conduct Survey

We considered 74 EDD services and software vendors for inclusion in our survey. From that initial set, we selected 18 which in our opinion would be most likely to provide us with the breadth and depth of data that we were seeking. Of that group, we were able to interview 14.

Starting with lists of the largest law firms and most prominent litigation departments and drawing on our own contacts, we selected 15 law firms to include in the survey. We eventually interviewed personnel from 7 law firms. The firms we interviewed were among the 250 largest in the country, ranged from international to regional in their market presence, had internal automated litigation support functions, and had been involved with electronic discovery in varying degrees.

We sought interviews with 11 corporate legal departments, conducted full interviews of 3 corporate legal departments and determined from short interviews with 2 other departments that they did not have enough EDD experience to respond to our questions.

While some interviewees were understandably guarded in their comments and cautious about divulging proprietary information, we saw an unexpected level of interest and enthusiasm and were gratified by the eagerness of many of the interviewees to share with us detailed and usually unobtainable information about their activities, structure, and even revenue.

Analysis

We entered interview information results into a spreadsheet constructed with a worksheet for each interviewee, and aggregated the individual responses into a master worksheet.

At the same time, we developed our list and definitions of market segments and refined the concepts. We defined 3 major segments and then sub-segments within them:

Computer forensics / investigative	
Data identification	Identifying what data should be pursued in discovery.
Data collection	Actual collection of data from various sources & media.
Data restoration	Restoring collected data for evaluation and processing.
Data conversion	Converting restored data to desired media & formats.
Culling / filtering	Culling: Reducing amount of data before processing. Filtering: Examining subsets of data after processing.

Electronic discovery / litigation support	
Data hosting / web	Hosting services of access to discovery data, including on-line hosting.
Filtering / search	Examining subsets of data; search, retrieval & display.
Data analysis	Working the legal team to discern significant patterns & located relevant documents.
Search software	Software provided to client for use in electronic discovery; litigation support.

Consulting	
Consulting	Advice on retention & discovery approaches.

We also created a set of 6 factors that formed the foundation of our evaluation of the information we had gathered, and assigned each factor a weight. The factors, weights and explanations were:

Factor	Weight	Explanation
Experience / reputation	20%	Length and depth of experience in the industry; reputation among customers and competitors.
Capacity	10%	Resources dedicated to the EDD market, such as facilities, people, and services.
Computer forensics / investigative	30%	Breadth and depth of services dedicated to forensic and electronic investigation.
Electronic discovery / litigation support	20%	Breadth and depth of services dedicated to conversion to more useable formats and to search, retrieval and analysis capabilities.
Consulting	5%	Advice on retention and discovery approaches.
Revenue	15%	Estimated annual revenues.

Within each of the main factors, we created a second level of factors. To these factors and sub-factors we assigned the questions and responses for analysis and synthesis. Next, we gave each of those questions a relative weight. The combined weight of questions within each sub-factor added up to 100%. The combined weight of questions within a factor added up to 100%. And, of course, the combined weight of all the questions within all the factors added up to 100%.

After defining factors and weighting questions, we rated the results for the 15 vendors about whom we had gathered information.

We then combined the rating and weighting information for each vendor to come up with composite numbers.

Synthesis

Composite Rankings

We developed composite numbers for 14 of the 15 EDD vendors. The numbers ranged from 3.7 up to 7.2:

Vendor rank	Composite number	Vendor rank	Composite number
1	7.2	8	5.5
2	6.9	9	5.4
3	6.5	10	5.4
4	6.1	11	5.3
5	5.8	12	4.9
6	5.6	13	4.0
7	5.5	14	3.7

The most important aspect of the composite numbers is their relative differences, not their absolute values. Of primary interest is the steadily descending slope from numbers 1 to 6, the compression from numbers 6 to 11, and then the sharp drop of the remaining vendors.

The composite rankings provide strong evidence in support of our subjective view that in this market there are many players but little differentiation between them, akin to early entrants in a campaign for political office.

This contrasts sharply with the view voiced by virtually every vendor we interviewed that each vendor believed it was one of the top 3 in the market.

Comparing the Top EDD Vendors

We compared the top 5 vendors using our 6 main factors – experience/reputation, capacity, computer forensics/investigative, electronic discovery/litigation support, consulting, and revenue. Of note was the variation – or lack of it – between vendors. The high-low ranges were:

Factor	High	Low
Experience/reputation	1.5	1.0
Capacity	.07	.05
Computer forensics/investigative	2.1	1.9
Electronic discovery/litigation support	1.3	1.1
Consulting	0.4	0.2
Revenue	1.4	0.8

Extending the comparison down to the top 8 vendors, we began to notice some trends. Although the vendors' capabilities (capacity, computer forensics/ investigative, and electronic discovery/litigation support) were almost identical in the aggregate, their presence in the market (experience/reputation) and revenue were not. The top three vendors enjoyed both the greatest presence and the largest revenues. Numbers 4 and 5 dropped off, and the remaining three vendors had numbers that were about half those of the leader.

Market Size – Past, Present and Future

Based on the information we were able to obtain through direct interviews and from secondary sources, noticeable expansion of the EDD market began in 1999. Starting at about \$40 million in revenues in 1999, the market appears to have grown to approximately \$70 million in 2000, and then \$150 million in 2001.

We estimated that the total 2002 domestic, commercial market for EDD services was at least \$270 million.

We project that for the next couple of years the EDD market will double annually. We anticipate, however, that a decrease in the rate of growth probably will happen starting soon after 2005, reflecting increased competition, technological and processing improvements, more judicial intervention and more targeted discovery demands, pressure to contain litigation costs, and better electronic records retention practices.

We want to stress that these numbers are both approximations and, we believe, conservative figures. For the most part, they do not include EDD services purchased by governmental units. The numbers are limited to activity within the United States even though there is EDD activity elsewhere. Finally, they are based in part on data from a limited number of EDD vendors. We suspect that information from a more in-depth survey would suggest larger figures.

The numbers also include figures for do-it-yourself players – law firms and corporate legal departments that have elected to conduct some or all of the EDD work internally. While this appears to be a small part of the market today, we anticipate that this segment of the market has the potential to grow considerably in the coming years and may dampen opportunities for vendors.

We also estimated the revenue of those vendors who were not willing to share that information directly, based on factors such as employee census, fee structures, processing methodologies, volume of material processed, and maximum project sizes.

Market Share

Our data indicates that no single vendor dominates the EDD market. With the same caveats mentioned above, according to our estimates the top 5 vendors commanded nearly 1/3 of the market, with shares ranging from just under 10% to a little less than 5%.

Market Perceptions

We asked all interviewees to identify the top 5 EDD vendors and to explain the reasons for their choices. Some interviewees provided numbered lists, others grouped vendors into tiers, yet others were unwilling to hazard any guesses. In responding, different interviewees focused on different criteria. Nonetheless, several criteria seemed to emerge on a regular basis. Among these, the most important seemed to be size, reputation, quality, and quantity and complexity of work undertaken.

Among the vendors we interviewed, 4 names dominated the lists; among the law firms, the first 3 names stood out. Those 4 vendors were 1-3 and 5 on our list.

Conclusion

As far as we know, this is the first study into the size, scope and growth of the EDD market where the authors of the study have been in a position to disclose the results. We anticipate that as we move forward, we will expand this study to include a larger group of vendors and respondents. We also plan to pursue related studies in areas such as the

factors that drive up and push down demand for EDD services, EDD vendor best practices, EDD consumer best practices, and the development and deployment of EDD standards.

To be continued....

The Authors

George Socha is an attorney and consultant who focuses on the effective use of technology to enhance the practice of law. Bringing to bear expertise with electronic discovery, data manipulation, and complex information management, he offers services as a special master, expert witness, and advisor. He can be reached at Socha Consulting LLC at 651-690-1739 or at Halleland Lewis Nilan Sipkins & Johnson at 612-204-4134 or gsocha@halleland.com.

Tom Gelbmann is an independent management consultant who has helped lawyers realize lasting value from information technology for over 16 years. He can be reached at Gelbmann & Associates at 651-483-0022 or tom@gelbmann.biz.