



Methodology

August 4, 2008

by

George J. Socha, Jr.
Socha Consulting LLC
651.690.1739
george@sochaconsulting.com
www.sochaconsulting.com

Thomas A. Gelbmann, Jr.
Gelbmann & Associates
651.483.0022
tom@gelbmann.biz
www.gelbmann.biz

I. Methodology

A. Research

1. Interviews and Spreadsheets

From December 2007 through May of 2008, we contacted, or were contacted by, more than 1,269 individuals about participation in our survey. Of the 1,269 we tracked, 784 individuals represented 372 providers of electronic discovery services or software, 342 individuals represented 216 law firms, and 143 individuals represented 95 corporations.

We also posted open invitations at www.sochaconsulting.com, on The Litigation Support Mailing List at litsupport@yahooogroups.com, and at various other venues.

We sent spreadsheets to 206 people who said they would like to participate in our survey during the earlier part of that period, and send additional spreadsheets later in the process. We sent each participant one of three spreadsheet survey instruments, depending on the participant's organization type – the law firm spreadsheet (Appendix 1), the corporate spreadsheet (Appendix 2), and the provider spreadsheet (Appendix 3).

We conducted telephonic interview of, or received spreadsheets back from, individuals at 155 organizations. Forty-eight of the organizations were consumers – 29 law firms and 19 corporations – and 107 were providers.

Some individuals chose to forego interviews, instead filling in the spreadsheets to the best of their ability and returning those to us. Most participants, however, elected to provide information through interviews supplemented by data entered into portions of the spreadsheets or sent via email.

We conducted the interviews over the phone. They lasted between 30 minutes and 2 hours. We usually we able complete the interviews in a single session; sometimes, however, the interview spanned several calls.

Following the completion of each interview, we sent a copy of the completed spreadsheet to the interviewee. We asked that the interviewee make any corrections or changes and either return the revised spreadsheet or send the updated information to us via email.

Almost all interviews were granted on the condition that we would ensure that it would not be possible for anyone to trace information back to the source. As a result of the need to provide this assurance, we ultimately concluded that we could publicly identify only those providers who came out of our analyses with the highest aggregate scores, and even then we could not tie individual names to individual scores or market shares. We also concluded that we could not disclose the identities of any of the consumers who gave us information.

2. Published Information

We gathered information from a variety of secondary sources as well. We collected data from websites for numerous organizations purporting to offer some form of electronic discovery services or software. We also drew on information available from a variety of business and information resources.

B. Analysis

1. Analytical Spreadsheets

We analyzed the data using the same fundamental approach we employed in previous surveys. Each spreadsheet consists of a number of worksheets, with each worksheet focusing on a specific topic such as “State of the Industry.” The company spreadsheet contains 35 worksheets to be completed in whole or in part; the law firm spreadsheet, 34 worksheets; and the provider spreadsheet, 39.

By the end of the information-gathering process, we have 155 spreadsheets containing information from or about 107 providers, 29 corporations and 19 law firms.

Once we finished collecting information, we normalized the data. By that we mean we took steps to ensure that the data was in a consistent structure and format so that we could do a better job of evaluating it. For example, we had dropdown menus for some items, such as states, but some people wrote over those menus, inserting “MN” for “Minnesota.” Other times, people placed content in a single spreadsheet cell where it ought to have gone into multiple cells, for example typing “10 gb” when they should have entered “10” in the “unit” cell and selected “Gigabyte” from dropdown menu in the next cell to the right.

Next we deconstructed the spreadsheets and reconstructed them by topic. For example, we copied the “State of the Industry” worksheet from every law firm spreadsheet we received or filled in as part of an interview, and pasted all those worksheets into a single new spreadsheet file, the law firm “State of the Industry” spreadsheet. We did the same for corporate and for provider “State of the Industry” worksheets. The consolidated spreadsheets served as the basis for our analysis. During this process, we created 146 consolidated spreadsheets.

For each consolidated spreadsheet, we prepared one or more summary worksheets the contained compilations of all the meaningful data contained in that spreadsheet’s worksheets. Using filters, PivotTables and a variety of functions, we prepared additional worksheets that served as the sources of information depicted in the charts in this report.

2. Ranking Models

We prepared additional spreadsheets specifically structured to allow us to rate and weight information from the consolidated spreadsheets, and from the rated and weighted information to prepare aggregated results. We constructed two models. We used one to rank providers of electronic discovery services, the other for providers of electronic discovery software. Were organizations offered both electronic discovery services and their own electronic discovery software, we included those organizations in both models.

Following the process we started last year, we gathered and analyzed data for individual stages of the electronic discovery process as defined in the Electronic Discovery Reference Model (EDRM). Those stages are Information Management, Identification, Preservation, Collection, Processing, Review, Analysis, Production, and Presentation.

We designed the spreadsheets to allow us to compare and cross-check data so that, for example, we could more readily identify inconsistencies between data points.

The ranking spreadsheet contains six levels of data. The top level consists of the major factors we considered as we evaluate services and software providers: Experience; software offered by providers; capacity; types of electronic discovery services; usage of software; law firm consumer rankings; corporate consumer rankings; and revenue. Levels two through six looked at increasingly specific details. For example, under "Capacity" (level one), we asked providers for the number of projects they handled (level two), how many of those were large projects as defined by each provider (level three), and of those how many involved collection of data (level four) (see Figure below). In all, we evaluated about 350 level one through level four categories (or rows) of information, with data coming from over 155 organizations.

Figure 1: Sample from Ranking Spreadsheet

1	2	A	B	C	D	E	F
-	103	100	Capacity				
•	104	101	No. of facilities/offices				
•	105	102	All offices				
•	106	103	EDRM functions				
•	107	104	Information Management				
•	108	105	Identification				
•	109	106	Preservation				
•	110	107	Collection				
•	111	108	Processing				
•	112	109	Review				
•	113	110	Analysis				
•	114	111	Production				
•	115	112	Presentation				

We assigned weights to each row of information. We entered data from the consolidated spreadsheets, assessed the array of data entered for each row, and assigned ratings to each item. We then aggregated the results so that ultimately we had a single composite score for each provider. Attached as an Appendix 5 is a listing of the level one through level three categories of information we considered, along with the weightings we assigned to those categories.

As we analyzed these numbers, we regularly checked with data from secondary sources to identify anomalies. Where necessary, we made corrections and then tested the corrections. In addition, we developed a series of additional models which enabled us to perform more detailed analyses of specific sets of information.

Consistent with the approach taken the last three years, we identified key factors around which we built our analysis of the data collected during the survey.

For the electronic discovery services model, the top-level factors and their weightings were:

- § Experience/Reputation (10%): We considered the length and depth of the providers' experience in the electronic discovery industry, the breadth and depth of the organizations' service offerings, and their reputations.
- § Capacity (15%): We looked at the level of resources that the providers dedicated to the electronic discovery market. Factors we considered included the number, experience and expertise of their personnel; the geographical coverage of their services; the number and configuration of its facilities; storage and processing capacity; and the number, size and duration of projects and types of data handled.
- § Types of Services (20%): We evaluated the types of services the providers offered as defined in by the Electronic Discovery Reference Model (EDRM).
- § Software Usage (3%): We looked at the breadth of software the providers used in delivery of services, including the level of software use in each stage of the electronic discovery process as defined by the EDRM.
- § Consumer Ranking – Law Firms (18%): This category focused on a variety of law firm perspectives about providers: Who the law firms used, who they preferred, and their perceptions of volume and quality of services and overall reputation.
- § Consumer Ranking – Corporations (18%): This category focused on the same perspectives from a corporate point of view: Who the corporations used, who they preferred, and their perceptions of volume and quality of services and overall reputation.
- § Revenue (16%): We took into account estimated annual revenues for the providers, calculated to exclude revenues for activities other than electronic discovery.

For the electronic discovery software model, we used a somewhat different set of factors and weightings:

- § Experience/Reputation (10%): We considered the length and depth of the providers' experience in the electronic discovery industry, the breadth and depth of their software offerings, and their reputation.
- § Software- Providers (3%): We looked at the breadth of use in the stages of the EDRM, intended installation (ASP and/or network environments), and frequency of use by law firms, corporations and electronic discovery service providers.
- § Capacity (5%): This category reflects level of resources that the providers dedicated to the electronic discovery market. Factors we considered included the number, experience and expertise of their personnel; the geographical coverage of their

products; types of data handled; and scalability of software tools.

- § Software Usage (30%): We look into account the breadth and level of use of the providers' software by law firms, corporations and electronic discovery service providers at each stage of the EDRM, as well as satisfaction ratings.
- § Consumer ranking – Law Firms (18%): This category focused on a variety of law firm perspectives about providers: Who the corporations used, who they preferred, and their perception of volume, quality of services and overall reputation.
- § Consumer ranking – Corporations (18%): This category focused on the same issues from a corporate point of view: Who the corporations used, who they preferred, and their perception of volume, quality, and overall reputation.
- § Revenue (16%): We took into account estimated annual revenues for the providers, calculated to exclude revenues for activities other than electronic discovery.

3. Market Sizing Model

Just as we have for the past five years, we constructed a model both to permit us to assess the likely size of the electronic discovery market in 2007 and to allow us to make market size projections for the next three years, 2008 through 2010.

We started with numbers given to us by participants – 2007 electronic discovery revenues reported by providers and 2007 electronic discovery expenditures reported to us by consumers. We added to that the providers' revenue expectations for 2008 through 2010 and the consumers anticipated expenditures for the same years. We also added provider and consumer expectations for market growth or shrinkage for the next three years. To that we added information collected both from participants and elsewhere about the probable number and size of electronic discovery providers and consumers in 2007. Finally, we drew upon numbers from the five previous surveys.

For our assessment of the size of the market for services and software in 2007, on the provider side we calculated five sets of numbers: total revenues, highest revenues for a single provider, lowest reported revenues, average, revenues, and median revenues. Working with those numbers and additional information, we arrived at three key figures for electronic discovery services: Tier 1 average revenue per organization, tier 2 average revenue per organization, and tier 3 average revenue per organization. Using analogous information from consumers, we arrived at a number to use for Do-It-Yourself Operations (DIYers) – consumers who elect to perform electronic discovery services themselves as opposed to sending the work to providers. Those numbers, in addition to software provider revenues combined to arrive at the 2007 market size estimate.

To project potential market growth or shrinkage over the next three years, we employed a series of calculations to arrive at four sets of growth or shrinkage numbers:



- § Maximum percentage expected growth or shrinkage for 2008, 2009 and 2010, as reported by consumers and providers;
- § Minimum percentage;
- § Median percentage;
- § Average percentage; and
- § Average percentage excluding maximum and minimum.

Using those figures and our assessment of the size of the market in 2007, we calculated potential market sizes for 2008 through 2010.

2008 Socha-Gelbmann Electronic Discovery Survey

Selected Items from Ranking Spreadsheet

Levels 1 through 3 of 6

Level 1	Level 2	Level 3	Weight 1 - Services	Weight 1 - Software	Weight 2 - Services	Weight 2 - Software	Weight 3 - Services	Weight 3 - Software
Experience			10.0%	10.0%	100.0%	100.0%		
	Length of time in industry				45.0%	35.0%	100.0%	100.0%
	Year established						20.0%	20.0%
	Year began ED						50.0%	40.0%
	ED services						30.0%	0.0%
	Year First Offered ED Software						0.0%	20.0%
	ED software - Year began						0.0%	20.0%
	Types of cases				10.0%	25.0%	100.0%	100.0%
	Types						50.0%	50.0%
	Number						50.0%	50.0%
	Who buys your ED services				40.0%	0.0%	100.0%	
	Law firms						40.0%	
	Corporations						40.0%	
	Governments						5.0%	
	Other service providers						15.0%	
	Who buys your ED software				0.0%	40.0%		100.0%
	Law firms							40.0%
	Corporations							40.0%
	Governments							5.0%
	Other service providers							15.0%
	Preferred provider				5.0%	0.0%		
Software - Providers' Own Use of Software			3.0%	3.0%	100.0%	100.0%		
	Number of programs used				33.3%	33.3%		
	Intended uses				33.3%	33.3%	100.0%	100.0%
	Information Management						12.0%	12.0%
	Identification						12.0%	12.0%
	Preservation						12.0%	12.0%
	Collection						12.0%	12.0%
	Processing						12.0%	12.0%
	Review						12.0%	12.0%
	Analysis						12.0%	12.0%

2008 Socha-Gelbmann Electronic Discovery Survey

Selected Items from Ranking Spreadsheet

Levels 1 through 3 of 6

Level 1	Level 2	Level 3	Weight 1 - Services	Weight 1 - Software	Weight 2 - Services	Weight 2 - Software	Weight 3 - Services	Weight 3 - Software
		Production					12.0%	12.0%
		Presentation					4.0%	4.0%
	Frequency of use				33.3%	33.3%		
Capacity			15.0%	5.0%	100.0%	100.0%		
	No. of facilities/offices				5.0%	5.0%	100.0%	100.0%
		All offices					40.0%	40.0%
		EDRM functions					60.0%	60.0%
	No. of employees				5.0%	5.0%		
	Geographical coverage - services				10.0%	0.0%	100.0%	
		International					25.0%	
		US					75.0%	
	Geographical coverage - software				0.0%	10.0%		100.0%
		International						25.0%
		US						75.0%
	Storage capacity				15.0%	15.0%	100.0%	100.0%
		Maximum					33.3%	33.3%
		Percentage of maximum used - average					33.3%	33.3%
		Percentage of maximum used - peak					33.3%	33.3%
	Processing capacity				20.0%	35.0%	100.0%	100.0%
		Maximum					33.3%	33.3%
		Percentage of maximum used - average					33.3%	33.3%
		Percentage of maximum used - peak					33.3%	33.3%
	Number of projects				15.0%	5.0%	100.0%	100.0%
		Overall					25.0%	25.0%
		Large					25.0%	25.0%
		Medium					25.0%	25.0%
		Small					25.0%	25.0%
	Types of data				30.0%	25.0%		
Types of ED Services			20.0%	0.0%	100.0%			
	ED Services - yes or no				50.0%		100.0%	
		Information Management					12.0%	

2008 Socha-Gelbmann Electronic Discovery Survey

Selected Items from Ranking Spreadsheet

Levels 1 through 3 of 6

Level 1	Level 2	Level 3	Weight 1 - Services	Weight 1 - Software	Weight 2 - Services	Weight 2 - Software	Weight 3 - Services	Weight 3 - Software
		Identification					12.0%	
		Preservation					12.0%	
		Collection					12.0%	
		Processing					12.0%	
		Review					12.0%	
		Analysis					12.0%	
		Production					12.0%	
		Presentation					4.0%	
	Year started services				25.0%		100.0%	
		Information Management					12.0%	
		Identification					12.0%	
		Preservation					12.0%	
		Collection					12.0%	
		Processing					12.0%	
		Review					12.0%	
		Analysis					12.0%	
		Production					12.0%	
		Presentation					4.0%	
	ED Services - %				25.0%		100.0%	
		Information Management					12.0%	
		Identification					12.0%	
		Preservation					12.0%	
		Collection					12.0%	
		Processing					12.0%	
		Review					12.0%	
		Analysis					12.0%	
		Production					12.0%	
		Presentation					4.0%	
	Software Usage - Usage of Providers' Software by Others		0.0%	30.0%			100.0%	
	Providers						33.3%	100.0%
		Program Usage						20.0%

2008 Socha-Gelbmann Electronic Discovery Survey

Selected Items from Ranking Spreadsheet

Levels 1 through 3 of 6

Level 1	Level 2	Level 3	Weight 1 - Services	Weight 1 - Software	Weight 2 - Services	Weight 2 - Software	Weight 3 - Services	Weight 3 - Software
		How Many Years Software Used						20.0%
		How Many Seats						20.0%
		How Frequently Used						20.0%
		How Satisfied						20.0%
	Law Firms					33.3%		100.0%
		Program Usage						20.0%
		How Many Years Software Used						20.0%
		How Many Seats						20.0%
		How Frequently Used						20.0%
		How Satisfied						20.0%
	Corporations					33.3%		100.0%
		Program Usage						20.0%
		How Many Years Software Used						20.0%
		How Many Seats						20.0%
		How Frequently Used						20.0%
		How Satisfied						20.0%
	Consumer Ranking - Law Firms		18.0%	18.0%	100.0%	100.0%		
	Services				50.0%	0.0%	100.0%	
		Who consumers used - services					50.0%	
		Who consumers preferred - services					50.0%	
	Software				0.0%	50.0%		100.0%
		Who consumers used - software						50.0%
		Who consumers preferred - software						50.0%
	Perceptions				50.0%	50.0%	100.0%	100.0%
		Consumer perceptions: volume					25.0%	25.0%
		Consumer perceptions: quality					25.0%	25.0%
		Consumer perceptions: reputation					25.0%	25.0%
		Consumer perceptions: project management					25.0%	25.0%
	Consumer Ranking - Corporations		18.0%	18.0%	100.0%	100.0%		
	Services				50.0%	0.0%	100.0%	
		Who consumers used - services					50.0%	

2008 Socha-Gelbmann Electronic Discovery Survey

Selected Items from Ranking Spreadsheet

Levels 1 through 3 of 6

Level 1	Level 2	Level 3	Weight 1 - Services	Weight 1 - Software	Weight 2 - Services	Weight 2 - Software	Weight 3 - Services	Weight 3 - Software
		Who consumers preferred - services					50.0%	
	Software				0.0%	50.0%		100.0%
		Who consumers used - software						50.0%
		Who consumers preferred - software						50.0%
	Perceptions				50.0%	50.0%	100.0%	100.0%
		Consumer perceptions: volume					25.0%	25.0%
		Consumer perceptions: quality					25.0%	25.0%
		Consumer perceptions: reputation					25.0%	25.0%
		Consumer perceptions: project management					25.0%	25.0%
Revenue			16.0%	16.0%	100.0%	100.0%		
	Services				100.0%			
	Software					100.0%		