eOPS 2010: Electronic Discovery Operational Parameters Survey

Executive Summary – April, 2010

Better information will make E-Discovery more efficient.

The multi-billion dollar electronic discovery market is not well understood. While effective initiatives to guide policy (such as the Sedona Conference1), working groups to collectively improve the mechanics of electronic discovery (the Electronic Discovery Reference Model2), and broad assessments of the economic scale of the market (the Socha-Gelbmann Survey3) have developed into valuable tools for buyers of electronic discovery technologies and services, there is no visibility into practical operational parameters across projects and across practitioners. Dutton, LLC provides electronic discovery management services, including purchasing decision support, workflow design, testimonial services and project management throughout the life-cycle of electronic discovery projects. From our vantage point, we are able to monitor interesting trends; however, most of our work cannot be disclosed due to various constraints, ranging from commercial non-disclosure agreements to court issued protective orders. Yet even with our unusual access to cross-project operational data, we have some questions. And we are not alone.

Participants at every level in the electronic discovery market have questions that, if answered, will help them improve efficiencies. Everything from purchasing decisions to resource planning during the review of potentially responsive documents can be improved with better information. What is “typical” for the volume of ESI per custodian? How many documents are typically in a GB? How many documents are being reviewed per reviewer per hour? How much do things cost? The lack of answers to these and many more questions make it difficult for consumers of electronic discovery services to make sound decisions and effective project plans.

1 www.thesedonaconference.org

2 www.edrm.net

3 www.sochaconsulting.com
The answers are not static. For example, both ESI volume and pricing change over time. As what is regularly referred to as “information explosion” continues unabated, it can be expected that more ESI will be collected, processed and reviewed per custodian over time. But how much more? And what is typical today? Pricing in E-Discovery has dropped significantly over time. How much?

To begin to answer, Dutton, LLC launched the first eOPS survey in March, 2010. This survey was established to gauge current parameters in electronic discovery that contribute to overall costs of collecting, culling, processing, reviewing and producing ESI in litigation that requires electronic discovery. We plan to re-survey the industry twice-yearly and look forward to being able to report trends from the baseline established by the results of this survey.

The mechanics of the survey are straightforward. We designed the survey questions and developed the survey on the on-line survey platform Zoomerang. To keep the administration of the survey simple and avoid potential issues associated with unsolicited email, we invited responses from firms we either work with directly, provide services to our clients, or have responded to RFPs we have managed for our clients; invited participants therefore are not random and do not necessarily represent the entirety of the electronic discovery business community. We believe the survey participants are sufficiently representative to provide the foundation for future trend analysis. The next pass of eOPS will be conducted in the fall of 2010 and will include a larger sample size.

4 http://www.zoomerang.com/Survey/WEB22ACBJNBSWD
The eOPS 2010 Respondents

Responses were collected between March 18, 2010 and April 2, 2010. Overall, 32% of the invited participants responded with at least partial information and 19% of the invited participants completed the survey. This is an extraordinarily high response rate.

Survey respondents are comprised of:

- Most (80%) of the responding firms have been in the electronic discovery business more than 5 years. Only 20% reported being in business from 1 to 5 years and no respondents have been in business less than a year. We interpret this as a maturing of the industry. Respondents were evenly split between large entities (40% have >500 employees) and small businesses (40% have 25-100 employees), with the remaining 20% reporting from 100 to 500 employees.

80% of respondents have been in business more than 5 years and 60% have more than 100 employees.
The Results
The survey was split into four sections:
1. Volume Parameters
2. Performance Parameters
3. Pricing Parameters
4. Trends

Volume Parameters
To get a sense of the scale of business of the respondents, we asked them what was the largest E-Discovery job (in terms of data volume) that they have been involved in during the past three months. While there is significant variability in the reported answers leading to a skewed mean (the maximum reported 400 Terabytes makes the mean unreliably indicate that large jobs are typical). Therefore, in this case, we believe the median is a better indicator of typical “large” jobs. The median reported was 3 Terabytes.

We also asked what the size (in Gigabytes) a “typical” E-Discovery job was during the past three months. Multiple Terabyte E-Discovery jobs remain the exception. Reported “typical size” E-Discovery matters had a mean of 476GB and median of 350GB.

While this self-reported scale is not an audit, it is clear that the respondents are accustomed to supporting significant E-Discovery projects.

Q: How much ESI is collected per custodian?
One of the operational parameters that drive the cost of electronic discovery is the volume of ESI collected per custodian. Collected ESI carries the cost of collection and drives the cost of culling, processing and review. Respondents reported that the volume of ESI (in GB) collected from various sources is:

<table>
<thead>
<tr>
<th>Source</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Sources</td>
<td>23</td>
<td>15</td>
</tr>
<tr>
<td>Email Servers</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Laptop/Desktop Hard Drives</td>
<td>22</td>
<td>18</td>
</tr>
<tr>
<td>File Servers</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
While it might be reasonable to anticipate that the sum of email server collections, hard drive collections and file server collections per custodian to add up to (or at least be less than) “all sources” per custodian, the data imply that there is inconsistency in the sources from which data are collected across projects. This rings true. For example, not all projects will include collections from all of the custodian’s hard drives, so the average volume of ESI collected from all sources can be less than the average volume collected from that particular source. Nonetheless, with “all sources” being less than “hard drives” it appears that there is significant opportunity to filter (or cull) ESI more aggressively from custodian laptop/desktop collections. Future versions of the survey will attempt to get at finer granular detail on this topic.

Q: What culling rates are being achieved?
Culling collected ESI is one of the most important means to manage the cost of review. Significant developments in culling technologies have been deployed in the market during recent years. We want to understand how effective they are. The reported culling rate (defined in the survey as the number of files loaded for review out of the number of files collected was consistent across respondents, with a reported mean of 67% and a median of 73%. While specific culling rates will always depend on the specifics of any matter, it is encouraging that culling rates are relatively high and consistent. We take this as evidence that as sophisticated culling methodologies are being developed and made available, they are being adopted. As information explosion continues at the custodian level, it will be even more important to implement effective culling strategies to ensure that reviewers are only asked to review potentially relevant ESI. We look forward to seeing how culling rates trend over time.

Q: How many documents are in a GB?
The average number of documents per GB is a significant variable in forecasting costs and resource requirements for the review phase of any E-Discovery project, since attorney review project managers typically measure documents coded per hour, rather than GBs coded, or pages coded, for example. Since there is a long tradition of measuring pages per GB (based on the old pricing model that charged per page delivered, rather than per Gigabyte processed), there is little data on how many documents are being extracted from collected Gigabytes for processing and review. We asked how many documents are in a GB of data, after culling. The mean of the reported average documents per GB is 5244 and the median is 5500.

Being able to predict the number of documents that need to be reviewed enables review teams to plan resources and respond effectively to scheduling orders.
Performance Parameters

We anticipate E-Discovery Performance will improve over time, as new technologies (software and infrastructure) and new methodologies (workflow refinements) are deployed. The Quality Control mantra applies here: “what is measured, improves.” In the E-Discovery market, performance in the service provider community has traditionally been focused on the units that drive the service providers’ revenue, such as Gigabytes. Whether it is Gigabytes culled, Gigabytes processed, Gigabytes loaded, Gigabytes stored per month, the performance units measured have been Gigabytes.

However, Gigabytes are not the units that drive overall E-Discovery project costs. The Gigabyte as a performance measure falls down in the review phase of a project, where the bulk of E-Discovery expense lives.

For that reasons, this first eOPS survey focused on one performance parameter: review throughput. The efficiency of review, in our experience, is the largest driver of overall cost of E-Discovery project. Indeed, in many matters, the cost of review is the single largest contributor to the cost of defense.

One of the trends in E-Discovery is the deployment of sophisticated software tools that allow review teams to make bulk-coding decisions on similar documents. Sometimes this takes the form of “near de-duplication,” sometimes it takes the form of variably selecting the “similarity” value to assemble clusters of documents for review. Such analytical tools can be powerful; however, they have to be used in ways that leverage that potential power. We found that 67% of the respondents are using analytical tools to automate at least some coding of documents and 78% of the respondents are using analytical tools in the design of their workflow.

The applications of analytical tools and other workflow design are only valuable if they influence the number of documents that can be reviewed per reviewer per hour. So in addition to asking whether analytical tools were applied, we asked:

Q: How many documents are reviewed per hour on your review team?

The review rate is the single most influential variable in the cost of an electronic discovery project. The respondents in this survey reported a mean of 84 and a median of 85 documents per reviewer per hour.
Pricing Parameters

Pricing in E-Discovery is notoriously oblique. It often seems like the balloon animal metaphor applies... when you squeeze one area, the air simply blows up another section. With no clear way to compare unit pricing across suppliers, pricing is sometimes confusing at best and misleading at worst. Having managed many millions of dollars worth of E-Discovery projects, we’ve selected key drivers in pricing that significantly drive the overall cost of an E-Discovery project. These key drivers are not intended to be comprehensive. This is not an analysis of all unit pricing across vendors, or even within a single vendor. Rather, they are some of the pricing elements we believe have significant impact on overall project cost. We asked respondents to report prices they paid or invoiced for these cost elements during the past three months.

Q: What is the average charged per hour for reviewers?
A: $104 (mean), $105 (median).

This reported cost per hour for review indicates significant reliance on contract review providers for first pass review among the respondents. The decision to use contract attorneys for review is a legal decision that must be made by counsel in each matter. It is not always appropriate and it is not always more efficient. This reported number, however, indicates that contract review is gaining acceptance, since typical in-house law firm review rates can be between $150 and $300 per hour. We expect that as the sample size in future surveys is expanded, this number may vary significantly depending on the makeup of the respondents.
Q: What is the average cost per GB for

<table>
<thead>
<tr>
<th>UNIT</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-review searching and indexing?</td>
<td>167.00</td>
<td>200.00</td>
</tr>
<tr>
<td>Processing (indexing, extracting and delivering to a review database)?</td>
<td>473.00</td>
<td>495.00</td>
</tr>
<tr>
<td>Hosting (per month)?</td>
<td>47.00</td>
<td>50.00</td>
</tr>
<tr>
<td>Analytical Tools (clustering, similarity analysis, etc.)?</td>
<td>304.00</td>
<td>150.00</td>
</tr>
</tbody>
</table>

This pricing information does not necessarily represent target price for any of these units. For any project, the overall forecast cost will depend on the characteristics of the legal requirements, the specifics of the client information systems, the total price package from the vendor, the experience of the review team and many other factors. These are, however, units that impact any E-Discovery project. If pricing on a specific project varies significantly from market pricing, it is an opportunity to engage in conversations about value.

To create effective cost projections for an E-Discovery project, E-Discovery buyers need to evaluate the total project cost including collection, preservation, processing, review and production. All costs, including attorney review costs, need to be included to evaluate the potential benefits of technologies and workflow enhancements. While certain technologies and methodologies may appear more expensive during the vendor assessment phase, the potential impact of each approach on the cost of review needs to be taken into account to reach sound conclusions.

When price points from vendors vary significantly from the mean market price, it presents an opportunity to engage in conversations about value.
Trends in E-Discovery

We asked respondents to report on what they believe are the most significant trends in E-Discovery. We appreciate the insights respondents provided. Some of the themes that emerged include:

1. Technologies that have an impact on the volume of data reviewed or on the efficiency of review workflow are emerging and will continue to gain acceptance. Examples include:
   - Early Case Assessment
   - “search-based” discovery
   - automating at least some of the coding/tagging that has traditionally been done exclusively by human reviewers

2. Pricing models that favor early culling
   - aggressive pricing of culling (one vendor offers free culling)
   - bundling of pricing, either per custodian or all-in per document reviewed

3. Shifting purchasing control
   - The trend toward E-Discovery purchasing decisions being made/influenced by corporations rather than being made exclusively by outside counsel continues to evolve.

At the highest level, knowing whether E-Discovery is a growing market or not is a critical piece of information for industry participants. While we believe the functional requirement for E-Discovery will continue, and the volume of ESI subject to discovery will continue to grow, it is not clear whether the services required to deliver that functionality will always require the same level of support from third party service providers. For example, it is possible that the promise of enterprise-wide Enterprise Content Management solutions will evolve to a point that they are useful in responding to comprehensive discovery requests.

So we asked our survey participants:

Q: During the next twelve months, will the market for electronic discovery services stay the same, shrink or grow?

There is consensus that the market for electronic discovery services is growing, but it is not a uniformly held belief.
Conclusions

This first eOPS Survey provides a baseline of useful information.

We believe there is significant evidence that the industry is maturing – the tools required to deliver efficient E-Discovery solutions throughout the life cycle of a project are being adopted and efficiencies are emerging. The idea of “best practices,” however, is still far away. There is not one “right” way, and the legal requirements of each matter continue to determine which, if any, technologies can be applied to automate various phases of E-Discovery.

Pricing is becoming more transparent. Armed with this transparency, E-Discovery consumers can engage in valuable conversations with service providers across the entire spectrum of E-Discovery projects and procure solutions that fit the specifics of their requirements more efficiently.

The market for E-Discovery services appears to be continuing to grow. However, fully a third of the respondents reported that they believe the market will either stay flat or shrink. If the market contracts, marginal providers may be especially hard hit.

The 2010 eOPS Survey provides a snapshot that we hope will be the baseline in a periodic survey of E-Discovery Operational Parameters that will further enhance transparency and contribute to efficiencies in the market.

About the Author
Cliff Dutton is founder of Dutton, LLC, a consulting firm that provides DISCOVERY+, the project management methodology that optimizes efficiencies throughout the electronic discovery project life cycle. Dutton, LLC is a member of the Chartis Executive Liability E-Discovery Consulting Panel. Mr. Dutton has designed and implemented systems that have processed billions of pages of evidence across thousands of legal, regulatory and administrative matters. He was the national technical editor for the Electronic Discovery Reference Model (EDRM-I). Cliff co-authored an analysis of the application of electronic discovery to corporate governance that was named one of the top-ten academic papers on corporate governance by the Social Science Research Network. He can be reached via email to cliff@duttonllc.com.