

eDJ Group

“Technology-Assisted Review” Series

The Market Overview

By: Greg Buckles, Barry Murphy, Mikki Tomlinson, and Jason Velasco

Table of Contents

Addressing The High Costs Of Legal Review	2
The Mechanisms of Technology-Assisted Review	4
A View From The Trenches	6
Case Study #1 – TAR Lets Lawyers Practice Law	8
Case Study #2 – The Need For Experience	10
Focusing In On Predictive Coding	11
Market Adoption Of Predictive Coding In 2012 Will Be Significant	11
Reducing Legal Review Costs Is The End Game	15
What Say The Courts?	19
Recommendations	20

■ Addressing The High Costs Of Legal Review

Organizations are justifiably concerned about the high cost of eDiscovery. In the last decade, companies have focused efforts on getting preservation and collection under control in order to avoid sanctions and what had been very high EDD processing fees. Recently, with processing fees coming back down to earth, focus has shifted again to the cost of legal review of documents. Most estimates peg legal review as somewhere between 60% and 70% of the total cost of eDiscovery.

Even in conservative estimates, legal document review can be a significant expense. Given the complexity of legal matters, diversity of organizations, and relative newness of digital evidence, there are few generalized statistics on eDiscovery costs. Consider, however, that it is not uncommon for large companies to get sued and have to collect upwards of 100 GB that are potentially responsive in a large case. Given typical assumptions about culling rates, amount of documents per GB, amount of documents reviewed per hour, and the hourly cost of review, these companies are looking at legal review costs of over \$5 million for each sizeable matter. In reality, many large companies have more than twenty discovery events per year. Now, think about the company that has just five significant discovery events in a year – the legal review costs are still staggering. It is no wonder why there is renewed focus on using technology to speed legal review and bring down costs.

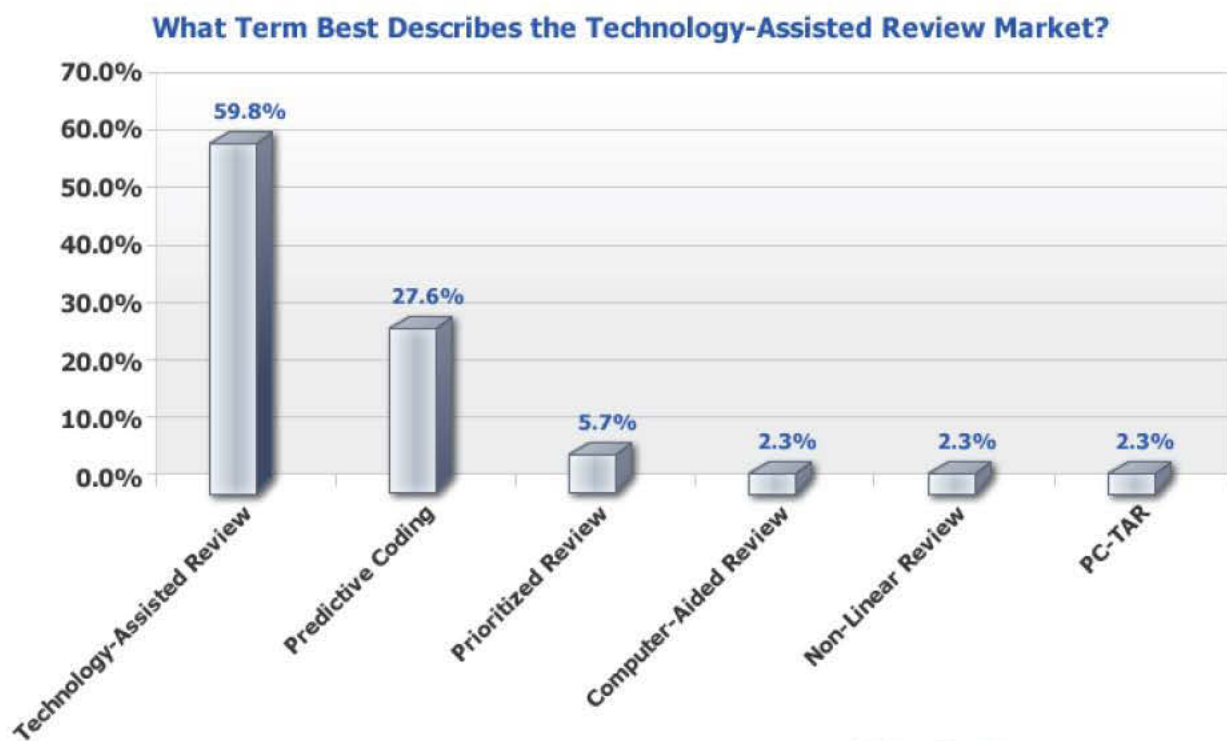
To assist organizations in cutting the costs of legal review, technology and service providers have come to market with a variety of what many call “technology-assisted review” – or TAR – offerings. In an effort to differentiate their offerings, vendors have introduced a variety of terms that actually confuse the market more than informing it. To date, eDJ has noted the following terminology to describe TAR market offerings:

- Predictive coding
- Meaning-based coding
- Adaptive coding
- Predictive Priority
- Transparent predictive coding
- Automated document review

¹ Assumptions: 100 GB of data with a culling rate of 65% results in 1,950,000 items for first-pass review; first-pass review with analytics at 250 documents per hour with hourly rate of \$75/hour; 30% of documents make it to second pass review where rates are 50 documents per hour at cost of \$400/hour.

Some argue that TAR is too broad a term for this market. However, terms like “predictive coding” are simply too narrow to cover the spectrum of this topic at hand. The bottom line is that, in the context of today’s advanced technological world, TAR is about using a combination of technology and people to actually speed, improve, and sometimes automate elements of the legal review process in such a way as to reduce costs and improve quality. Recent studies suggest that there is “evidence that such technology-assisted processes [those that use automated tools to prioritize and select documents for review], while indeed more efficient, can also yield results superior to those of exhaustive manual review, as measured by recall and precision, as well as F1 [which is] a summary measure combining both recall and precision.” To say that TAR can yield superior results does not necessarily mean that every variety of TAR will – companies using it will need to marry the right tools, workflow, and human expertise in order to create a defensible and measurable TAR process.

Figure 1: What Do We Call This Market?



Source: eDiscoveryJournal Poll, February - March 2012, N=87



² Maura R. Grossman & Gordon V. Cormack, Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review, XVII RICH. J.L. & TECH. 11 (2011), <http://jolt.richmond.edu/v17i3/article11.pdf>.

Given the confusion over the many names in the market, eDJ conducted a poll on The eDiscovery Journal throughout the months of February and March of 2012. Almost 60% of respondents prefer the broader term Technology-Assisted Review.

As a result, eDJ will now use Technology-Assisted Review (TAR) to describe this market.

The Mechanisms of Technology-Assisted Review

There are three main mechanisms, or methods, for using technology to make legal review faster, less costly, and generally smarter.

- **Rules-driven.** “I know what I am looking for and how to profile it.” In this scenario, a case team creates a set of criteria, or rules, for document review and builds what is essentially a coding manual. The rules are fed into the tool for execution on the document set. For example, one rule might be to “redact for privilege any time XYZ term appears and add the term ‘redacted’ where the data was removed.” This rule-driven approach requires iteration to truly be effective. The case team will likely have rules changes and improvements as the case goes on and more is learned about strategy and merit. This approach assumes that the case team knows the document set well and can apply very specific rules to the corpus in a reasonable fashion.
- **Facet-driven.** “I let the system show me the profile groups first.” In this scenario, a tool analyzes documents for potential items of interest or groups potentially similar items together so that reviewers can begin applying decisions. Reviewers typically utilize visual analytics that guide them through the process and take them to prioritized documents. This mechanism can also be called present-and-direct.
- **Propagation-based.** “I start making decisions and the system looks for similar-related items.” This type of TAR is about passing along, or propagating, what is known based on a sample set of documents to the rest of the documents in a corpus. In the market, this is often referred to as predictive coding because the system predicts whether documents will be responsive or privileged based on how other documents were coded by the review team. Propagation-based TAR comes in different flavors, but all involve an element of machine learning. In some scenarios, a review team will have access to a “seed set” of documents that the team codes and then feeds into the system. The system then mimics the action of the review team as it codes the remainder of the corpus. In other scenarios, there is not a seed set; rather, the systems give reviewers random documents for coding and then create a model for relevance and non-relevance. It is important to note that propagation-based TAR goes beyond simple mimicry – it is about creating a linguistic mathematical model

for what relevance looks like.

These TAR mechanisms are not mutually exclusive. In fact, combining the mechanisms together can help overcome the limitations of individual approaches. For example, if a document corpus is not rich (e.g. does not have a high enough percentage of relevant documents), it can be hard to create a seed set that will be a good training set for the propagation-based system. However, it is possible to use facet-based TAR, for example concept searching, to more quickly find the documents that are relevant so as to create a model for relevance that the propagation-based system can leverage.

It is important to be aware that these approaches require more than just technology. It is critical to have the right people in place to support the technology and the workflow required to conduct TAR. Organizations looking to exercise these mechanisms of TAR will need:

- **Experts in the right tools and information retrieval.** Software is an important part of TAR. The team executing TAR will need someone that can program the tool set with the rules necessary for the system to intelligently mark documents. Furthermore, information retrieval is a science unto itself, blending linguistics, statistics, and computer science. Anyone practicing TAR will need the right team of experts to ensure a defensible and measurable process.
- **Legal review team.** While much of the chatter around TAR centers on its ability to cut lawyers out of the review process, the reality is that the legal review team will become more important than ever. The quality and consistency of the decisions this team makes will determine the effectiveness that any tool can have in applying those decisions to a document set.
- **Auditor.** Much of the defensibility and acceptability of TAR mechanisms will rely on the statistics behind how certain the organization can be that the output of the TAR system matches the input specification. Accurate measures of performance are important not only at the end of the TAR process, but also throughout the process in order to understand where efforts need to be focused in the next cycle or iteration. Anyone involved in setting or performing measurements should be trained in statistics.

For an organization to use a propagated approach, in addition to people it may need a “seed” set of known documents. Some systems use random samples to create seed sets while others enable users to supply small sets from the early case investigations. These documents are reviewed by the legal review team and marked as relevant, privileged, etc. Then, the solution can learn from the seed set and apply what it learns to a larger collection of documents. Often this seed set is not available, or the seed set does not have enough positive data to be statistically useful. From what eDJ hears from professionals using TAR, the practice has value, but

requires a sophisticated team of users (with expertise in information retrieval, statistics, and law) who understand the potential limitations and danger of false confidence that can arise from improper use. For example, using a propagation-based approach with a seed set of documents can have issues when less than 10% of the seed set documents are positive for relevance. In contrast, rules driven and other systems can result in false negative decisions when based on narrow custodian example sets.

However TAR approaches and tools are used, they will only be effective if usage is anchored in a thought out, methodically sound process. This requires a definition of what to look for, searching for items that meet that definition, measuring results, and then refining those results on the basis of the measured results. Such an end-to-end plan will help to decide what methods and tools should be used in a given case.

A View From The Trenches

Sometimes, the most valuable way to learn about a topic is to hear the stories of those actually practicing it. To that end, eDJ interviewed several users of TAR. Some lessons learned from real-world users include:

- Iteration is necessary. Users will be constantly training and re-training the system as they learn new facts about a matter.
- TAR is both an art and a science. Reviewers need to be creative about ways to reduce the volume of information to review. At the same time, they need to be able to apply science (in this case, statistics and technology) to the process.
- Using TAR can allow law firms to actually charge a higher per document fee for review because it is possible to both reduce the volume of documents reviewed while increasing the speed of review.
- It is possible to ease into using TAR without necessarily automating the review process and skipping exhaustive manual review altogether right away. One idea is to use various TAR techniques for prioritizing documents for review and for quality control (QC). As users become more comfortable with the tools, techniques, and process, usage can evolve further into culling and building the review sets.
- More targeted collections – and therefore higher relevance rates – would make TAR even more effective. Often, lawyers are conservative and cast a wide collection net resulting in low relevance rates that make it difficult to get a good seed set. Another good practice is to add a culling step post-collection (or at the point of collection) to remove junk (e.g. ESPN newsletters), which will only serve to increase relevance rates in the

remaining documents.

- The makeup of a document collection has a large bearing on whether the results of TAR will be effective or not. If data sets are bilingual, for example, text analysis may not work as well as hoped, depending on the technology in use. Some products are able to automatically group documents based on language, allowing users to create training sets for each language and design workflows that can be effective. This is an example of how combining a facet-based approach with a propagation-based approach can lead to better, faster, and more accurate review. The larger point here is that effectiveness of predictive coding will ultimately be dependent on the makeup of the document collection and users should be aware that challenges will arise when collections have diverse data such as bilingual content or images that cannot be OCR'd.
- New forms of data, such as social media, may pose challenges for TAR. There is no way of doing text analysis on things like Tweets because of the abbreviations and short words in this new form of communication.

Case Study #1 – TAR Lets Lawyers Practice Law

A law firm we spoke with described the key driver to use TAR – in this case, predictive coding – was the clear message from corporate clients that the cost of document review is too high and too unpredictable. In looking at how to impact the cost of review, the firm boiled it down to three variables:

1. Cost of the reviewer
2. Speed of review
3. Volume of documents

At the time, there was a tendency to look at document review as a commodity because it was possible to lower the cost of the reviewer by outsourcing (either internationally or to lower-cost American cities). But, outsourcing only serves to distance the attorneys in charge of strategy from the facts. This firm wanted to review documents in a less expensive way, but also in a way that improved quality; that extra distance from the facts seemed contrary to that goal. The firm realized that by utilizing advanced technology, it is possible to lower the cost of review while honing in on case facts more quickly. And, it is possible to do that in a single process and a single platform – that reduces complexity.

The firm began the process of moving into predictive coding with the underlying knowledge that it was simply time to replace the document review tool that had been in place. A predictive coding tool was chosen in order to address the two variables of cost the firm wanted to impact: speed of review and volume of documents. Because this particular firm is committed to alternative fee engagements and because clients complain about the unpredictable nature of review costs, predictive coding proved to be even more valuable because the technology allows more accurate predictions of review budgets.

The firm manages the full review process, including the technology and the legal reviewers, internally. There is a team that specifically handles the technology involved with predictive coding, a robust staff attorney team, and an eDiscovery attorney team. In general, the predictive coding technology is user-friendly and straightforward, but the firm stresses the need to really know how to use it. Since this firm owns the tool and employs the reviewers directly, its attorneys get more and more experience with the tool. Additionally, the firm has invested in significant training on how predictive coding workflows operate and advanced statistics.

For this firm, predictive coding is more than simply marking some documents responsive and then training the system. Truly, it is about understanding the way the technology works and how the algorithm works and then

getting commonality amongst the documents being worked on in order to understand what the training set needs to consist of. The key is to know how the system judges commonality amongst documents. For example, a matter may have keywords that would not necessarily indicate commonality, such as “baseball” and “stock.” The firm learned to first train the document set for baseball and then train the set for stock.

Commonality is what the predictive coding technology hangs its hat on. There needs to be some commonality amongst the documents. The technology is looking at ways in which words appear in a document in the same context – looking for documents with similar characteristics. Thus, the firm learned that, in this use case, predictive coding is better at helping to code responsive documents than it is at coding things non-responsive. This is because the review team knows generally what makes a document responsive as opposed to what makes a document non-responsive.

One major statement this firm made is that predictive coding is an art and a science and that it is crucial to use a combination of TAR methods if necessary to help chip away at the volume of documents. It requires experience, facility with the technology, and the ability to be smart and creative to understand what can be eliminated to get the document volume down to a small universe and then use predictive coding to get through those documents faster.

Another key point is that predictive coding is not static. Iteration can be very helpful. For example, it is possible to take 5 documents out of 100 and use those to train the system. That might find 20 other documents that are similar to the original 5. If you add a custodian and more documents added, you can just apply the work done already as the “seed” set. The training is not necessarily instantaneous, but it does not take long. With the technology this firm uses, all of that work happens in the background while the tool is in use. It is important to consider, however, that adding documents to a seed set changes probabilities and does present some risk to rolling deliveries.

For this firm, predictive coding has been a win-win. Clients are happy because of cost savings. There have been matters that have allowed the firm to save half of what the cost would have been without predictive coding. For the firm, the per document cost goes up because it can employ higher-value reviewers while only reviewing a fraction of what it would have had to previously. The firm also benefits by using less and less contract attorneys. Perhaps the best result is that the firm remarked, “clients are coming to us for legal advice as opposed to document review.” At the end of the day, litigators want to litigate.

Case Study #2 – The Need For Experience

One firm we spoke with tested predictive coding to see whether the upfront cost of the software would pay off with downstream cost savings. The firm had a previously reviewed case with 500K documents. Using keyword filters, the corpus was reduced to 100K, of which about 3K records that were relevant and produced.

Starting over, and using the predictive coding mechanism, the firm sought to see if better results were possible. The lead lawyer from the review team spent a few days training the system and then the 500K documents were run through the solution. The first issue occurred immediately; the relevance rate (a document being marked as positive for relevance) in the set of documents was so low in the corpus that it became difficult to train the system with any degree of statistical significance. The lesson learned was that, for predictive coding to be effective, the seed set of documents needs a certain percentage of documents to be positive for relevance. This firm believes that the seed set needs at least 10% of documents to be positive in order to be effective. Some systems combine a known relevant seed set with random sampling to tackle this issue. All of this highlights the higher level of expertise required to leverage these new technologies.

The firm tested predictive coding again under different conditions. In the next case, there were 80K documents in the total set from a user's workstation; keyword search and file type filters brought the corpus down to 21K documents. After the legal team conducted review there were 800 documents found to be relevant. To test predictive coding, the firm used the set of 21K documents as the seed set, telling the system that the 800 documents were positive for relevance. The predictive coding solution was then applied to the rest of the 60K documents in the total corpus. The results showed that there was some inconsistency between what the lawyers had deemed relevant and what the predictive coding solution ranked high for relevance. There were documents that were found to be relevant that the predictive coding solution ranked low, and some documents that were marked non-relevant that the predictive coding solution ranked high (technically, they were relevant, but they were almost exact duplicates with different hashes; like a word file with a tiny different bit of metadata). During the manual review, one of the instructions to the reviewers was to only mark one copy as relevant (because the firm was using near-duplicate identification). This likely happened because the near duplicate documents were marked non-relevant when seeding the predictive coding solution and that marking confused the solution. The lesson learned here is that any organization using predictive coding has to be very careful to understand the seed document set, how that set should be programmed into the solution, and how the statistics work so that the practice can be defensible. There are many complexities to the process that can derail an inexperienced organization.

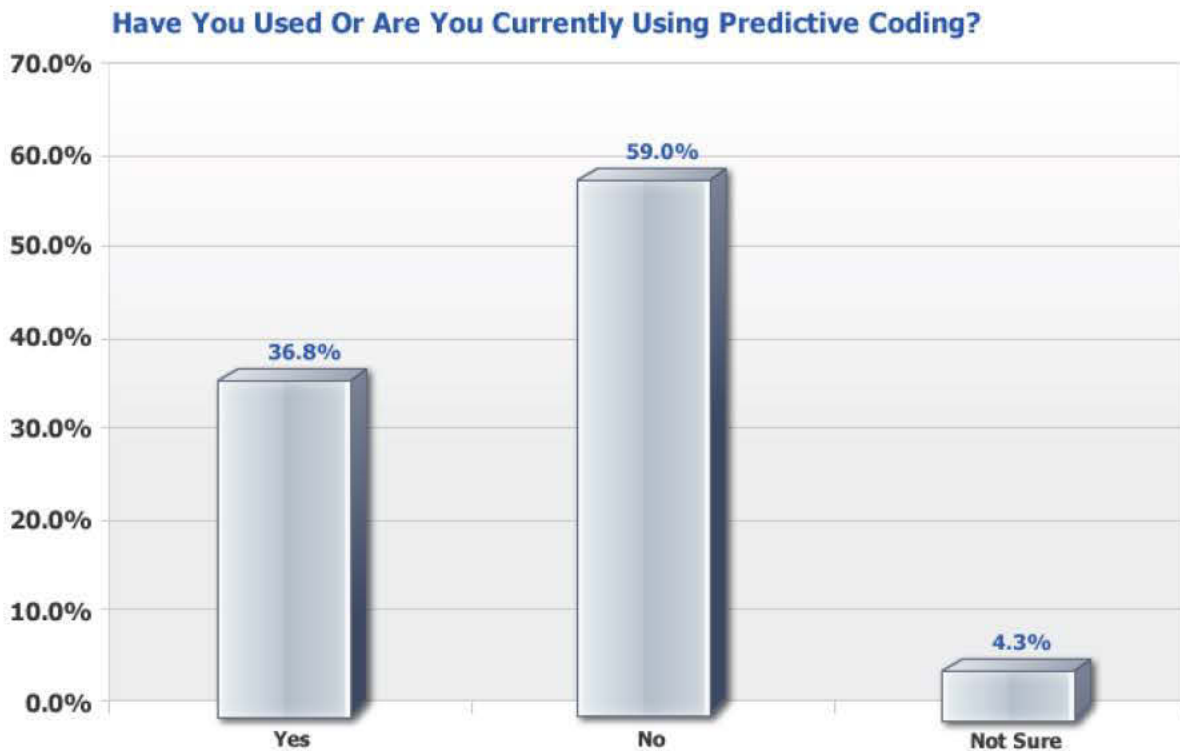
Focusing In On Predictive Coding

In early 2012, eDJ Group conducted a survey of 117 Legal IT professionals to understand what eDiscovery practitioners think about one specific type of TAR – predictive coding. In that survey, eDJ inquired about how much experience they have with predictive coding, what plans for future usage are, and what the expected benefits and concerns are.

Market Adoption Of Predictive Coding In 2012 Will Be Significant

As is often the case, buzz about a topic pre-dates the topic actually being a mainstream practice. In the case of predictive coding, this is certainly true. A large majority of respondents – 59% - indicated that they have not yet utilized predictive coding.

Figure 2: Predictive Coding Gaining Traction



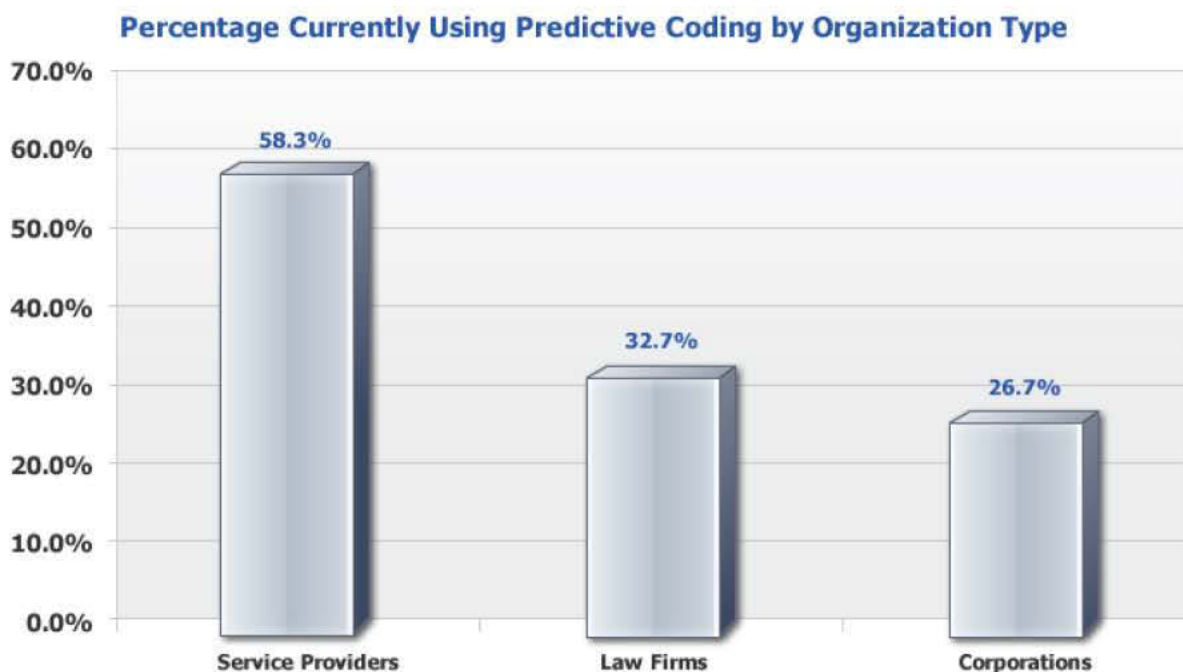
Source: eDJ Technology-Assisted Review / Predictive Coding Survey, March 2012, N = 117



Drilling a bit further into that adoption figure, law firm and corporate adoption of predictive coding is lower than the general market number. A full 70% of corporations and 60% of law firms have not yet experimented with predictive coding. Meanwhile, 58% of service providers have adopted predictive coding. This is logical when

considering how important people are to the practice. Service providers have expertise on staff that can manage predictive coding projects. These providers also need to be a step ahead of their customers – corporations and law firms – in order to be ready once those organizations demand more predictive coding expertise.

Figure 3: Service Providers Leading the Way in Adopting Predictive Coding



Source: eDJ Technology-Assisted Review / Predictive Coding Survey, March 2012, N = 117

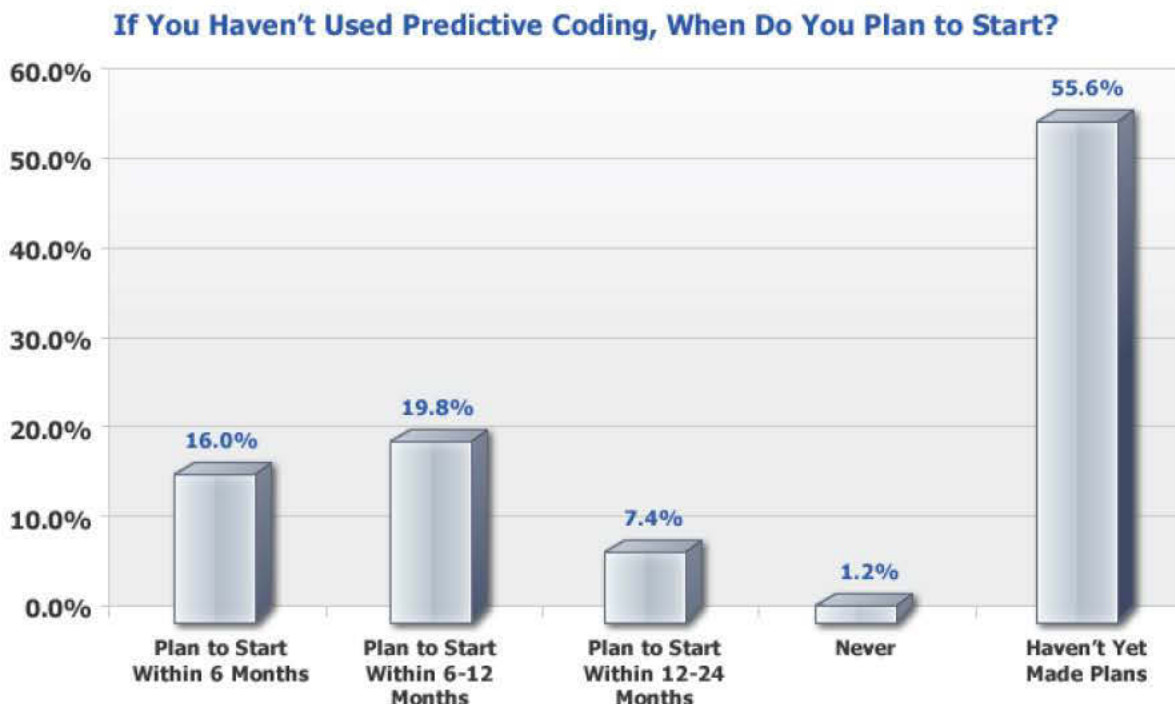


Do not think, however, that the buzz around this topic is unwarranted. That almost 37% of respondents have used or are using predictive coding is significant. The legal community tends to adopt new solutions in a slower, more conservative manner given that new approaches need to be battled-tested and court-ready. That over a third of respondents are already in the trenches indicates that advanced technology-assisted review methods are poised to take off.

How quickly will the market take off? eDJ believes usage of predictive coding will actually take hold more quickly than one might think. Because the legal market tends to be conservative, technology and solutions markets within the legal realm tend to grow more slowly than other mainstream markets, e.g. customer relationship management (CRM). In this case, however, the legal market shows signs of being ready to accept predictive coding sooner than later. Of those respondents that have not yet used predictive coding, 35.8% plan to start before 2012 is over. By

the end of 2012, almost 62% of respondents will have adopted predictive coding, giving the practice significant market penetration.

Figure 4: Usage of Predictive Coding Will Grow, But Many Still Uncertain

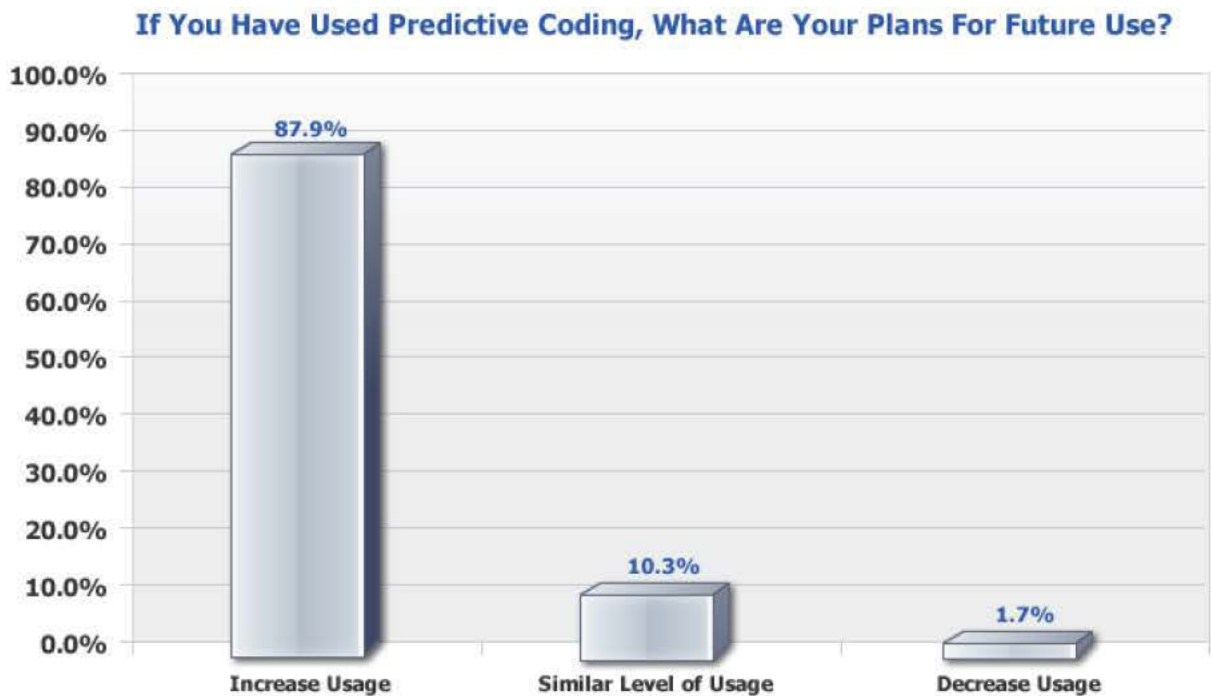


Source: eDJ Technology-Assisted Review / Predictive Coding Survey, March 2012, N = 81



The market for predictive coding will grow not only from the addition of new users, but also from increased usage by existing users. Of those that already use predictive coding, almost 90% plan to increase usage going forward. Clearly, users are getting value by leveraging predictive coding.

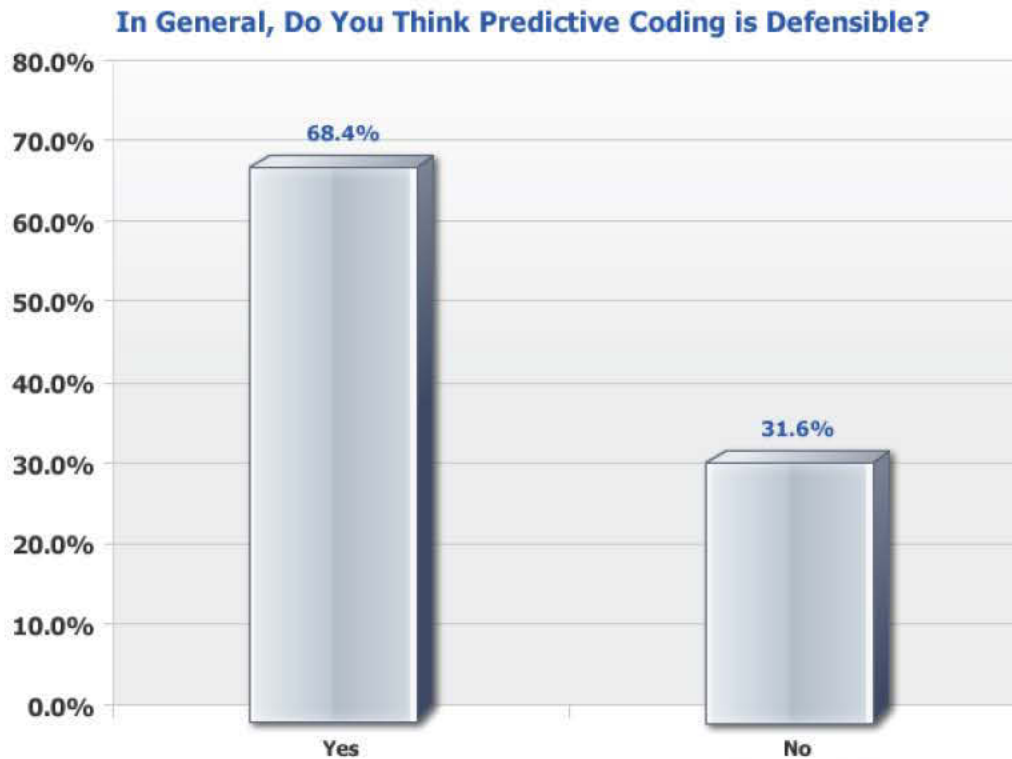
Figure 5: Users Getting Value Out Of Predictive Coding



Source: eDJ Technology-Assisted Review / Predictive Coding Survey, March 2012, N = 58

Paving the way for broad market adoption is a general acceptance of the defensibility of predictive coding. Beginning in February 2010, eDiscoveryJournal.com ran an informal poll asking readers' opinions on the defensibility of predictive coding. During the first few months the poll was live, only about 25% of readers thought predictive coding was defensible, but that percentage began to grow – and grow rapidly – as time went on. Respondents to eDJ's formal survey on the topic tend to believe that, in general, predictive coding is defensible. Less than one-third indicated that they think predictive coding is not defensible.

Figure 6: General Belief is that Predictive Coding is Defensible



Source: eDJ Technology-Assisted Review / Predictive Coding Survey, March 2012, N = 117



■ Reducing Legal Review Costs Is The End Game

Predictive coding can provide multiple benefits, but across the board, respondents from both corporations and law firms rank cost reduction as the primary driver.

Figure 7: Cost Reduction is the Main Driver for Predictive Coding

Rank the Following Predictive Coding Drivers in Order of Importance to You.

Corporation	
1	Cost
2	Improving Accuracy / Quality
3	Review Time
4	True Early Case Assessment - Settlement
5	Managing Risk / Exposure Mitigation
6	Enablement of Small Internal Review Teams; Elimination of Outside Review
7	Reduction of Collection / Preservation Volumes

Law Firm	
1	Cost
2	Review Time
3	Improving Accuracy / Quality
4	True Early Case Assessment - Settlement
5	Managing Risk / Exposure Mitigation
6	Reduction of Collection / Preservation Volumes
7	Client Pressure to Utilize

Respondents were asked to rank order the value drivers for predictive coding. The rankings here are a weighted average.



Source: eDJ Technology-Assisted Review / Predictive Coding Survey, March 2012, N = 117

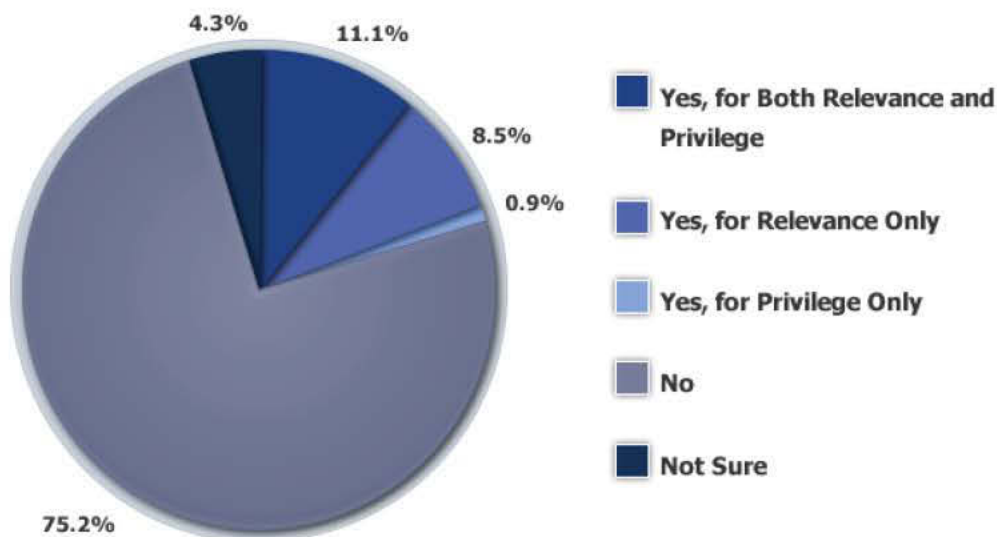
For law firms, cutting review time is also very important because law firms must ensure that resources are used as efficiently as possible in order to maintain reasonable margins. Although it is a common belief that law firms have a vested interest in maximizing associate review time, firms that eDJ has spoken with want to maximize legal decisions and practicing law rather than trudge through millions of documents. For corporations, minimizing review time is also important – more so because lower review times translate into lower review costs. Both corporations and law firms value the ability to improve the accuracy and quality of review in general. As previously mentioned, studies have shown that TAR can achieve more accurate results than manual human review. After all, humans can be inconsistent in how they judge various documents for responsiveness and/or privilege. With technology-assisted review, it is possible to combine the subject-matter expertise of human reviewers with technology and processes that bring speed and consistency to legal review.

Also ranking high as a driver for predictive coding is true early case assessment – the ability to quickly understand the major elements of a matter and make the right decisions about the case based on merit. In fact, this early decision component of technology-assisted review is the one most cited to eDJ analysts anecdotally. Those using

techniques like predictive coding are doing so to make merit-based decisions more quickly. While predictive coding could potentially eliminate the need for Counsel to individually review each document for relevance and/or privilege calls, that is not what is happening in the real world today. Less than one-quarter of respondents have ever relied on predictive coding to make actual calls on documents without Counsel's review.

Figure 8: Majority Still Have Counsel Review Documents for Relevance/Prfivilege Calls

Have You Ever Relied On Predictive Coding To Make Actual Relevance Or Privilege Calls On Documents Without Counsel Individually Reviewing The Documents?

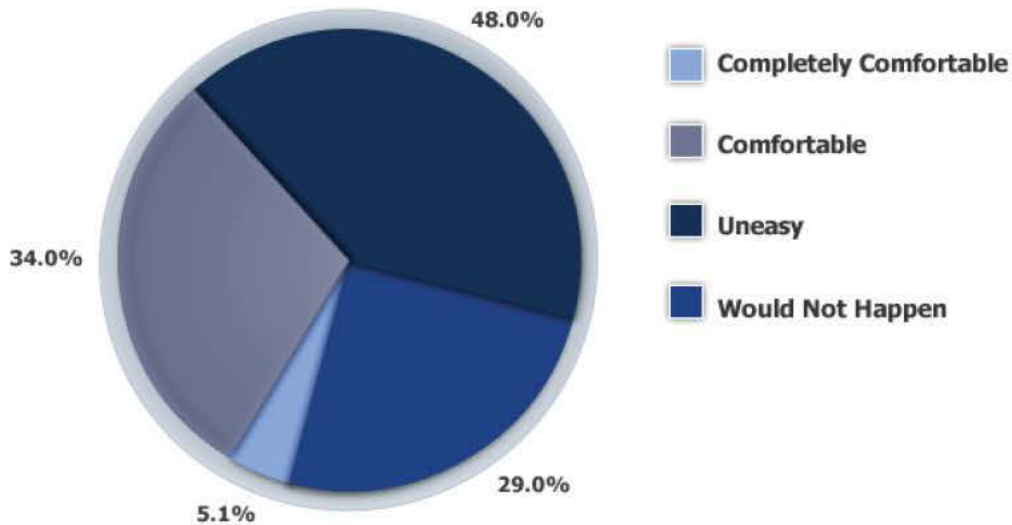


Source: eDJ Technology-Assisted Review / Predictive Coding Survey, March 2012, N = 117

This raises the question of whether techniques like predictive coding could spell the end of linear review as we know it. The answer is yes, but it is not likely to happen in the short-term. Almost one-third of respondents indicate that going directly to production without conducting linear review simply would not happen. And another 48% would not be comfortable with such a scenario.

Figure 9: There Is Uneasiness With Skipping Linear Review Completely

How Comfortable Would You Be With Using Predictive Coding To Go Directly To Production Without Conducting Linear Review?



Source: eDJ Technology-Assisted Review / Predictive Coding Survey, March 2012, N = 117

It is interesting to note, however, that 39% of respondents would be comfortable with skipping linear review and going straight to production. This means the market is evolving and attitudes are beginning to shift. As more and more cases gain attention and the benefits of technology-assisted review are proven again and again, eDiscovery professionals will become more comfortable with skipping linear review in the right situations.

While one might think that technology-assisted review will lead to a decrease in legal jobs and the elimination of human reviewers, survey respondents see through that myth. Most – 62.4% - do not see predictive coding as a threat to legal work or jobs. Certainly, there are those that are threatened or that are not yet sure if predictive coding is a threat, but in general, eDiscovery professionals believe that these advanced review methods are focused on making reviewers better at high-value work rather than at eliminating the need for reviewers.

Figure 10: Most Do Not View Predictive Coding As A Threat To Legal Jobs



Source: eDJ Technology-Assisted Review / Predictive Coding Survey, March 2012, N = 117



■ What Say The Courts?

Prior to February 2011, there was not a prescriptive law or case precedent that approved or mandated the use of TAR or describes the steps and methods one should use. But, on February 24, 2012, The Honorable Judge Andrew J. Peck did issue an opinion in the Da Silva Moore case, saying, “this judicial opinion now recognizes that computer-assisted review is an acceptable way to search for relevant ESI in appropriate cases.”

eDJ does not view this opinion as carte-blanche approval of TAR; rather, it is an official recognition by Judge Peck that, when utilized correctly, TAR can be an effective option for finding relevant information (as of the publication

³ Monique Da Silva Moore vs. Publicis Groupe. Case 11 Civ. 1279. United States District Court Southern District of New York. February 8, 2012.

of this report, the opinion is under appeal). The Judge had already written publicly a similar viewpoint, saying, “[I]f the use of predictive coding is challenged in a case before me, I will want to know what was done and why that produced defensible results. I may be less interested in the science behind the ‘black box’ of the vendor’s software than in whether it produced responsive documents with reasonably high recall and high precision... Proof of a valid ‘process,’ including quality control testing, also will be important.”

This harkens back to the point that effective TAR requires more than just a black box software tool. A transparent process is necessary in order to document how decisions are made. And, the right people are critical: an expert review team to make early relevance decisions; technical experts to set up software tools; and statisticians to ensure that the output of the system matches the rules of input with the right confidence levels.

■ Recommendations

Reducing and better managing the costs of legal review imperative for all corporations and law firms alike. Technology-assisted review can help, but the mechanisms and tools and services that support them are still evolving. Some courts are also very accepting of more automated review techniques as long as they are defensible and agreed up on by parties in advance. If considering a technology-assisted review project, you should:

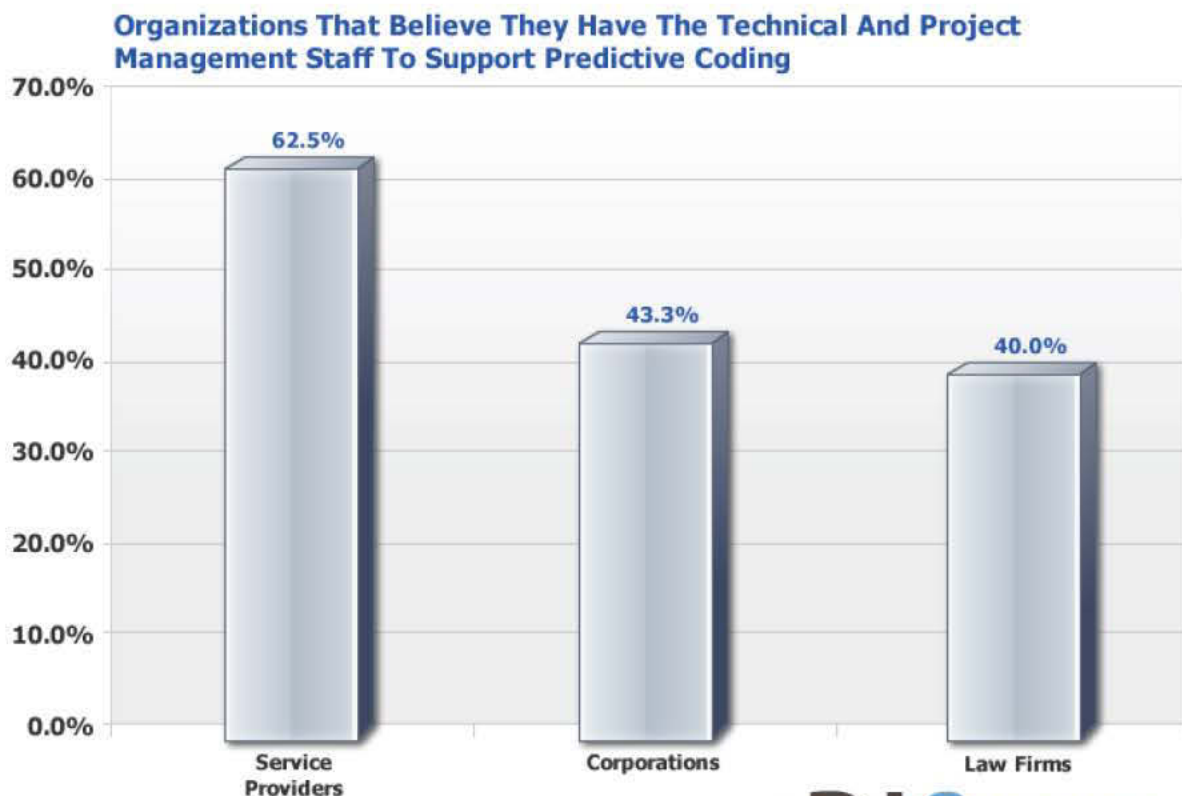
- **Consider the people factor.** Does your organization have the internal expertise to conduct technology-assisted review in a defensible manner? Many times, using a service provider can make sense due to the people factor alone. Less than half of the corporations and law firms in eDJ’s survey feel as though they have the technical and project management staff to support technology-assisted review.
- **Consider the different skill sets involved in TAR.** Beyond having a good legal review team (after all, the results of techniques like predictive coding are only as good as the quality of the seed set), it will be necessary to have staff skilled in the new technologies for TAR as well as statisticians that can validate how the output of the system matches the input.
- **Combine different TAR techniques together.** Users do not need to pick one form of TAR and simply run with it. For example, it is possible to use keyword searching to do initial culling, then use conceptual

⁴ Peck, Andrew. Search, Forward. Legal Technology News. October 1, 2011.

clustering to group records and do further culling, and use that culled set as the training sample for a propagation-based approach.

- **Beware of a non-standardized solutions market.** The marketplace will present many solutions with sexy names, all claiming to do similar things. The onus will be on the consumer to understand what the offerings are, how they are enabled technically, and how they are deployed. Offerings will come from technology vendors and service providers alike. Consumer should be very skeptical of any solution presented as a black box, where you can simply push a button and get results. TAR involves people and process in addition to technology, and it is inherently complex.

Figure 11: Human Resources Are An Important Consideration



Source: eDJ Technology-Assisted Review / Predictive Coding Survey, March 2012, N = 117



- **Communicate with other parties early on about the usage of TAR.** Early communication is the key to gaining the opposition and court's acceptance of TAR usage. The legal profession does not like to admit that linear human review is not perfect on very large collections. Explaining the goals, methods and limitations of TAR beforehand limits the instinctive reaction to attack the process in our adversarial system. While you may not go as far as the recent Da Silva Moore matter and agree that the opposition can participate in the

TAR training process, early metrics and sample production sets are strategies that can engage the other side. The courts are not going to fight a TAR solution when both parties agree on it.

In the longer term, the technologies and processes that drive TAR will likely be able to extend to broader information governance activities. eDiscovery professionals need to remain pragmatic, however. TAR can pay off today by decreasing legal review costs. It is best to experiment with solutions like TAR on specific projects where it makes sense. Then, take what is learned on those projects, optimize it, and apply to other projects as needed. The practice of TAR will grow considerably in 2012, and case law and best practices will begin to evolve.

■ About the eDJ Group

eDJ Group offers unbiased information and pragmatic advice, based on years of experience and proven industry best practices. Whether researching a technology or service solution, conducting an eDiscovery Bootcamp or finding the right expertise to answer your specific questions, eDJ Group is the source for all eDiscovery professionals.

We are committed to helping eDiscovery professionals get the information necessary to excel in their professions, rather than offering legal advice or counsel. We operate with the utmost integrity and commitment to our clients on these guiding principles:

- **Independence** – All research, reports, advice and services are agnostic and conducted independently without influence by sponsors.
- **Highest Ethical standards** – All content is honest perspective based on real experience and interactions with thousands of practitioners; detailing both successes and failures without favoritism.
- **Pragmatic, Experienced Expertise** – All services are conducted by industry experts with decades of experience in eDiscovery and strictly vetted by the eDJ Group founders.

For further information about the eDJ Group and their research, please contact Barry Murphy (barry@edjgrouping.com) or Jason Velasco (jason@edjgroupinc.com).