



Analytic Adoption: 2014 Market Report

eDJ Group Research Report

By:

Greg Buckles, Founding Consultant

© 2014, EDJ Group, Inc. All rights reserved. Customers that bought this report may make one attributed copy or slide of each figure contained herein. Licensed for reproduction by Acaveo, additional reproduction is strictly prohibited. Information is based on best available resources. Opinions herein reflect judgment at the time of research and are subject to change. Report illegal copies of this report to info@eDJGroupInc.com to receive a free copy plus another research report of your choice.

About the eDJ Group Inc.

The eDJ Group Inc. offers expert perspective, unbiased information, and pragmatic advice on eDiscovery and Information Governance products, companies, technologies and best practices. eDJ Group consultants have over 20 years professional experience in forensics, litigation, corporate governance, software design and many more related fields.

About the Author



Greg Buckles is an independent eDiscovery consultant specializing in enterprise technology and work flow solutions with over 25 years in discovery and consulting. His career spans law enforcement, legal service provider, corporate legal, law firm and legal software development. This deep and diverse background combines with exposure to the discovery challenges of Fortune 500 clients to provide a unique industry perspective.

Disclaimer:

EDJ Group is not a law firm. All expressed opinions and content are provided for general educational purposes only and are not specific legal advice, even if the author is a practicing attorney. Neither eDJ Group Inc. nor the information contained herein should be used as a substitute for competent legal advice from a licensed professional attorney in your state.

EDJ Group believes reasonable efforts have been made to ensure the accuracy of all eDJ Group Inc. original content. Content may include inaccuracies or typographical errors and may be changed or updated without notice. All eDJ Group original content is provided "AS IS" and while we endeavor to keep the information up to date and correct, we make no representations or warranties of any kind, express or implied, about the fitness for a particular purpose, completeness, accuracy, reliability, suitability, or availability with respect to the information, products, services, or related graphics for any specific purpose. Any reliance you place on such information is therefore strictly at your own risk.

In no event will EDJ Group or any of its contributors be liable for any direct, indirect, punitive, incidental, special, or consequential damages or damages for loss of profits, revenue, data, down time, or use, arising out of or in any way connected with the use of the document or performance of any services, whether based on contract, tort, negligence, strict liability or otherwise.

Contents

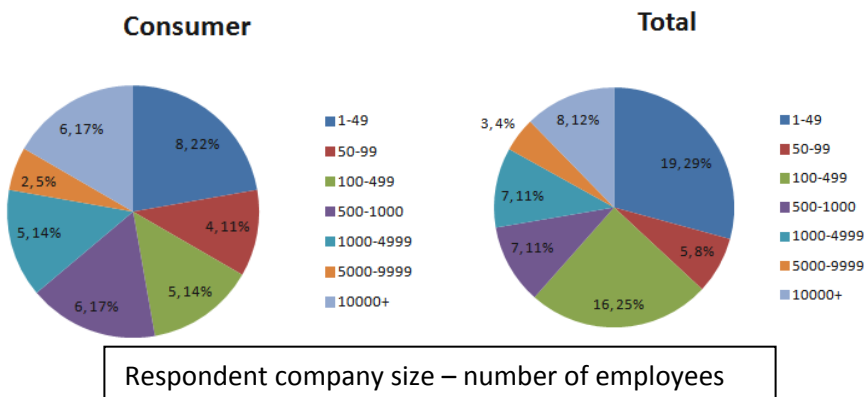
Overview and Purpose	4
Research Methodology	4
Defining Analytics	5
History	5
Review Approaches.....	6
PC/TAR Training Methods.....	7
Analytic Usage Cases.....	8
Analytics Market – eDiscovery & IG.....	9
Consumers	9
Buying Categories	10
Consumption Models.....	10
Survey Metrics – A Limited View	11
Consumer Interview Perspectives	13
Corporate Perspective	13
Law Firm Perspectives.....	14
Technology Agnostic Provider Perspective.....	15
eDJ Perspective	16
PC/TAR Adoption Rate	16
PC/TAR Adoption Challenges	17
The Sweet Spot of the Maturity Curve	19
Analytic Market.....	20
Executive Summary.....	21
Appendix 1: eDJ Matrix Analytic Offerings	22
Appendix 2: Aggregate Survey Responses.....	23
Appendix 3: New eDJ Matrix Features and Descriptions.....	26

Overview and Purpose

Predictive Coding (PC), Technology Assisted Review (TAR), Auto Categorization, Social Networking, Chronological Analysis and now even Topic Modeling¹ are all based on data engines that identify and extract patterns, exceptions and visualizations from large collections of relatively similar data sources. The market has long lumped all of these usage cases under the Analytics² label, starting back with the earliest enhanced review offerings from Attenex and Stratify back in 2001 and 2006, respectively. Jason R. Baron and Bennett B. Borden wrote a good law review article³ that lays out perspective on how analytics have crept into legal practice. The flood of PC/TAR articles, caselaw, blogs, case studies and other marketing hype has raised the question, “How are major law firms and corporations REALLY using these new toys?” The eDJ Group could not find any independent market research on actual analytic adoption rates and practices, so we conducted our own surveys and interviews to get some solid metrics to share with our consulting clients and all of you Participating Members⁴ who have taken an eDJ Group survey. The goal of this research cycle was to understand analytic usage cases, consumption models and the customer pain points that justify the potential additional expense. The eDJ Group has created new analytics market categories for the eDJ Matrix to help buyers with their own solution design. The research identified at least 36 technology offerings with some kind of clustering, machine learning, data visualizations or other analytic functionality.

Research Methodology

Surveys: The initial eDJ Group Analytic Adoption Survey was limited to consumers (n=38), but a second survey for eDiscovery providers (n=33) was added in response to early provider interviews. The surveys consisted of seven questions that could be interpreted independently or as aggregate metrics (combined n=71). The surveys ran on the eDJ Group website for roughly three months and participants were restricted to validated eDJ Group site members. This increased the quality of the responses while reducing the volume. As you can see below, the consumer respondents were heavily weighted towards large corporations (>1,000 employees) while eDiscovery providers averaged (<500 employees).



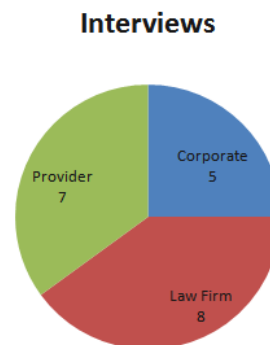
¹ <http://blog.nuix.com/2014/09/17/textual-analytics-topic-modeling/>

² <http://en.wikipedia.org/wiki/Analytics>

³ <http://jolt.richmond.edu/index.php/finding-the-signal-in-the-noise-information-governance-analytics-and-the-future-of-legal-practice/>

⁴ <http://www.edjgroupinc.com/page/faq>

Interviews: eDJ Group conducted 20 phone and live interviews that averaged 60-90 minutes. Interviews reviewed the individual survey responses to gather more information and validate (or invalidate) the value of the question. Although most phone interviews were scheduled for only 15 minutes, the respondents were eager to discuss their usage cases and the challenges they faced from customers and senior decision makers when proposing the use of machine learning for actual review decisions. This theme will be explored in detail within this report. All interview respondents had 10+ years of eDiscovery experience and had managed or conducted reviews using PC/TAR analytics. eDJ Group deliberately solicited interviews with respondents having deep expertise, market perspective and reputations as early adopters with heavy discovery/regulatory burdens.



Analytic Provider Briefings: Early responses to Survey Question 7⁵ enabled eDJ Group to identify analytics providers with the highest level of brand awareness. eDJ Group conducted briefings with these brand leaders and other providers with strongly differentiated usage cases or functionality, for a total of 11 briefings. The briefings focused on critical analytic functionality, market adoption and forward looking innovation trends. eDJ Group generally conducts annual product briefings for providers with significant market share or differentiating innovation that justify the briefing time. The objective features and subjective eDJ Notes (consultant perspectives) for all analytics providers are available to Participating Members in the eDJ Matrix⁶.

Defining Analytics

History

Following the first waves of email intensive litigation in the late 1990’s, Attenex and Stratify brought advanced analytics to the review of native files. These systems enhanced linear review by clustering related items for consistent, faster review at a relatively high price point. At this same time, many sophisticated litigants and providers used Boolean searches, advanced sorting and subject matter expertise to manually reorganize large collections to manage the cost of review. All of these methods leveraged relatively transparent technology to enhance the quality and limit the quantity of review with the goal of reducing review costs. The underlying assumption was that every item would get some kind of human review. The lack of clear rules or caselaw made most counsel uncomfortable with “Black Box” systems that promised to liberate them from eyes-on review of every item. Now that the federal and state judiciary has begun to accept some method of PC/TAR as a practical necessity in the face of ballooning ESI collection volumes, counsel and clients are struggling to let go of the “eyes on every item” assumption and make the risk/cost decision on every case with large discovery. At the same time, innovative new offerings seek to shortcut the entire relevance review with visual storyboarding,

⁵ Appendix 2: Aggregate Survey Responses

⁶ <http://www.edjgroupinc.com/>

interactive profiling and other investigative analytics focused at identification of key evidence well before traditional review phase.

Review Approaches

eDJ Group's research and surveys made it abundantly clear that 'Analytics' means many different things to the broad eDiscovery market. Even terms like Predictive Coding or Machine Learning can involve an infinite number of combinations of technologies, sampling strategies, training iterations and much more. Every provider seems to have a different take on how to leverage analytics before and during review. So what flavor of PC/TAR is right for you? For most consumers, the first priority is to differentiate between the major approaches that are offered as alternatives to traditional linear review.

Prioritization/Acceleration – Before review, the ESI is organized or clustered by a wide variety of manual or automated methods to improve review speed and decision consistency. Although reviewers can bulk mark clusters/stacks/threads, the general market assumption is that all items have had at least one set of eyes on them or another document so similar as to make no difference. The technique was introduced to the market over 10 years ago by Attenex, Stratify, Cataphora and others. Providers report up to 300% improvement in review speed depending upon the collection composition. This approach is considered fairly conservative and has low adoption resistance. Interviews indicated that most mature litigants are using some form of optimization prior to review. Techniques such as email threading, near duplicate clustering, conversation aggregation and topic/custodial clustering have been used for years.

Propagation – Reviewer decisions from seed and training sets are used to train a relevance algorithm on a single or multiple issue basis. That 'engine' classifies unreviewed documents into training categories (relevant, non-relevant, privileged, etc). Some systems use random training samples while others select samples across the categories/clusters for what some providers call "active learning". Everyone seems to have their own 'secret sauce' in sampling strategies. The propagation approach offers the highest potential review savings, but also faces the highest resistance according to interviews. Biglaw firms and many providers of contract reviewers perceive these systems as a direct threat to their revenue. Too many of these systems are considered 'black box' technologies, requiring a subject matter expert to validate, operate or explain. The latest generation have worked hard to simplify and visualize the feedback metrics that the user uses to make the decision on when the training process can stop. You will hear providers talk about stability, confidence, recall, precision and other concepts for measuring the effectiveness of the system.

Recommendation – This approach uses the same kinds of training algorithms as propagation systems and frequently prioritizes review batches based on document similarity, shared concepts and other data characteristics. The system displays issue recommendations and sometimes weights to assist the reviewer, but the final decision is made by a person. You could consider this a dynamic machine learning feedback system to improve review speed and consistency. Since the system itself is not applying review decisions, it may face less resistance for new users.

Quality Control/Pattern Analysis – The goal of these systems is to improve the quality and consistency of the review rather than decrease the cost. The engine uses similarity, concepts, word clusters and

more to compare the decision patterns and spot potential false negatives/positives. This is especially valuable in matters with heavy privilege or confidentiality issues.

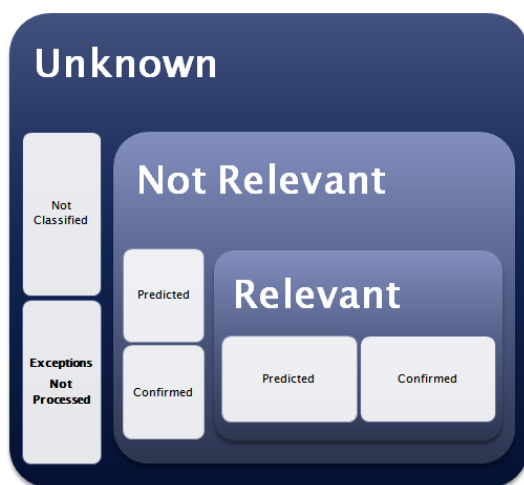
Ranked Navigation – This functionality can be used to identify the key documents from an opposing production, internal investigation or other early case assessment scenario. The solution interface and workflow enables a user to identify a relatively small number of the most important documents from the collection through clusters, sampling, profiles, social networking and other features.

PC/TAR Training Methods

Machine Learning⁷ should be renamed to “Machine Training” as the ‘machine’ learns from iterative review sets and builds a relevance algorithm/profile until it reaches a preset level of ‘stability’, usually expressed as a 95-99% Confidence Level. This report will not delve into the detailed terminology, sampling approaches or pros/cons of different systems. Instead, it will attempt to introduce this esoteric art to a new buyer contemplating and trying to differentiate multiple offerings.

Training Set Composition – How does the process select training documents?

- **Random Samples** – Can play a role in QA/QC, establishing the relative proportion of relevant documents (richness) and avoiding the trap of only finding the topics that you already know about.
- **Seed/Known Relevant Samples** – Using key documents from custodial interviews or other sources .
- **Predicted Relevant Samples** – Documents that the system already thinks are relevant. Some systems use continuously select and present the highest relevant items until a threshold of non-relevance is reached. All items produced will be reviewed. This method can reduce the volume of non-relevant documents inadvertently produced or exposed to demands for access to training documents.
- **Unclassified Samples** – Focuses on documents that do not match the either the relevant or not relevant profiles, but they contain enough text to be processed.



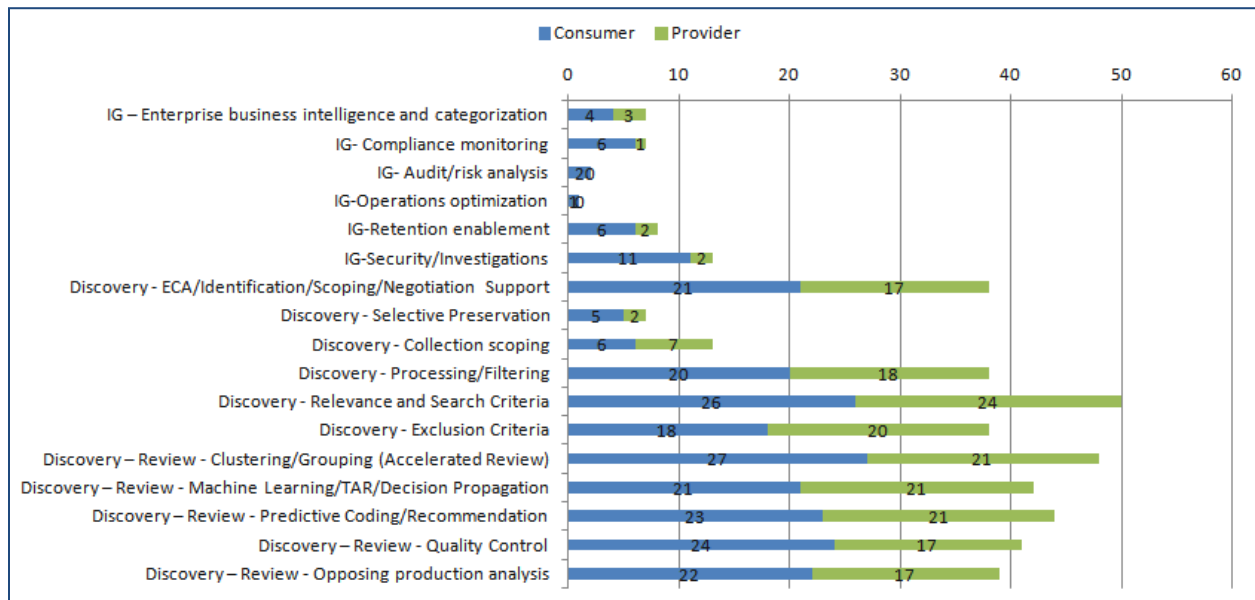
Size/Number of Training Passes – How many documents per training pass and what is the expected number of training passes given the collection size and composition? eDJ Group encountered extremely wide variations in the recommended/required training pass sample sizes (40-5,000) and expected numbers (10-40). Consistent best practice was to establish the estimated relevance richness (percentage relevant documents) prior to determining the training process or completion criteria.

⁷ http://en.wikipedia.org/wiki/Machine_learning

Active vs. Passive Learning/Sampling – All the marketing hype and blog wars about whose PC/TAR is more effective seem to be fighting over a very small number of customers who have already decided to use machine learning to tackle very large matters. Recently branded terms such as ‘Continuous Active Learning’ (CAL), ‘Simple Active Learning’ (SAL) or ‘Simple Passive Learning’ (SPL) can be confusing. Most providers use ‘Active’ to indicate that their process dynamically incorporates iterative training decisions when selecting new training sets. So ‘Passive’ refers to random statistical sampling that does not factor in predicted relevance. More important to most buying decisions is whether the training process can accommodate multiple relevance categories and rolling collections.

Analytic Usage Cases

Analytics have a wide variety of potential usage cases for Information Governance (IG upstream) and eDiscovery (downstream). The combined eDJ Group survey results indicate that corporations are just starting to evaluate enterprise analytics. The interviews and briefings gave better insight into the survey metrics.



Survey Question 3: What general usage cases to you/your customer use analytics on?

Usage Case	#	%	eDJ Perspective/Comment
IG – Enterprise business intelligence and categorization	7	1.60%	Large enterprise players such as IBM, HP/Autonomy and SAP have dominated Big Data business intelligence with structured database analytics. eD players such as Equivio and Nuix are just moving into this area with initial offerings. Interviews indicate that only the very early adopters are experimenting with these systems so far.
IG- Compliance monitoring	7	1.60%	Endpoint security and Data Loss Prevention providers have begun to add analytics on email and SharePoint. The key roadblock is dedicated resources to create and maintain monitoring rules and escalation workflow.
IG- Audit/risk analysis	2	0.46%	This is a reactive usage case that is frequently done in an ad hoc manner. Given the high probability that confirmation of malfeasance or fraudulent actions will result in subsequent regulatory or litigation disclosures, these investigation analytics should have full audit trail and decision tracking to validate potential evidence and minimize risk of spoliation.
IG-Operations optimization	1	0.23%	Large enterprise providers have been pitching these operations center dashboards and visual reporting for years without substantial success on

			the sales cycle. They seem to be a 'nice to have' but not a deciding feature on purchasing decisions.
IG-Retention enablement	8	1.83%	Enterprise archives have tried to leverage analytics to assign retention categories for at least 10 years with few real success stories outside of the financial vertical with regulated retention of all communications. eDJ is cautiously hopeful about recent case studies with reduced effort to train initial machine learning engine and more 'real world' workflow promise.
IG-Security/Investigations	13	2.97%	The strongest corporate IT buyers of analytics is the Security and Compliance departments to use in reactive investigation scenarios. These cases tend to have strong executive backing and can access funding with quick board approval. This reminds eDJ of the early discovery fire drills during the 2000-2005 period. The strong upturn in regulatory requests observed this year provides further motivation for corporation governance bodies to understand their risk and potential liability using the latest technology.
eD - ECA/Identification/Scoping/Negotiation Support	38	8.70%	16% of consumer survey respondents indicated that the primary benefit of analytics was for potential strategic advantages. This justifies ECA and scoping use of analytics on early collections.
eD - Selective Preservation	7	1.60%	eDJ has observed serious challenges to clients who want to apply holds selectively. Most do not have the staff expertise or manpower to perform defensible scoping workflows and settle for custodial or even departmental holds in the early days prior to actual discovery demands.
eD - Collection scoping	13	2.97%	Once the discovery demand has been negotiated or received without agreement, eDJ is hearing slow progress from more mature corporate legal departments leveraging analytics to define actual collection criteria. This is still a tough sale to most retained counsel, who prefer to just collect/process the broad custodial ESI.
eD - Processing/Filtering	38	8.70%	All of these usage cases fall under the concept of leveraging analytics to better shape and organize broad collections prior to review. Interviews confirmed that counsel accepted processing such as near deduplication, conversation aggregation and other relatively untested functionality that exclude ESI prior to review. This is mostly a game of semantics where counsel can say that they reviewed ALL processed ESI while excluding 30-50% based on rules and subjective criteria judgments. It will be interesting to see how this plays out when opposing counsel identify key evidence excluded in processing.
eD - Relevance and Search Criteria	50	11.44%	
eD - Exclusion Criteria	38	8.70%	
eD - Review - Clustering/Grouping (Accelerated Review)	48	10.98%	
eD - Review - Machine Learning/TAR/Decision Propagation	42	9.61%	As you will see in the following areas, although many litigants have used PC/TAR for a few matters, there is significant resistance to wholesale use of the technology to assign relevance to processed documents.
eD - Review - Predictive Coding/Recommendation	44	10.07%	
eD - Review - Quality Control	41	9.38%	This usage case is driven by review providers who are highly conscious of the risks rather than corporate legal departments.
eD - Review - Opposing production analysis	39	8.92%	Interview respondents cited heavy use of analytics on opposing productions because of impending deadlines and difficulty in justifying full review costs.

Analytics Market – eDiscovery & IG

Consumers

Consumers of analytics comprised many different market segments with unique pain points, feature requirements and even consumption models. Marketing and sales directors abhor a complex market space because of the challenges it presents to create a simple, clean brand and value proposition for their offerings. Thus the annual cycle of eDiscovery buzzwords splashed across LTN conference booth banners and LTN ads. However, oversimplification and high pressure sales of the latest PC/TAR variation can easily lead to buyer's remorse. This has resulted in eDJ client requests to break down the analytics market(s) and players into something easier to understand. That brings us back to the actual who and why of analytic buyers in four initial segments: **Corporate, Firm and Provider**. This report does not cover the fourth buying category, Software Providers, though the research did cover the OEM/API market place where analytic engines are integrated into other software offerings.

These consumer segments each align with some of the usage scenarios discussed above. However, the functional requirements and workflow for collection scoping by a corporate legal department on live enterprise sources may differ dramatically from those of a service provider profiling early documents from key custodians. Even when the technology features line up, the different consumers may require distinct licensing, pricing and delivery models.

Buying Categories

Corporate Buyers – Corporate buyers seek to proactively (IG) and reactively (eDiscovery) manage the volume of enterprise data in place. They want analytic functionality that can broadly profile, categorize and extract relevant data facets to enable defensible deletion, smart migration, selective preservation, relevance scoping, investigations and ECA scenarios on primary unstructured enterprise data sources. Although there is a market for collection/processing platforms that enable corporate litigation support to cut out the \$45-150/GB processing fees by providers, this can be considered a legacy buying scenario. Recent eDJ RFP support engagements have all shifted to direct inventory, search and analysis of native data in place over broad preservation collections.

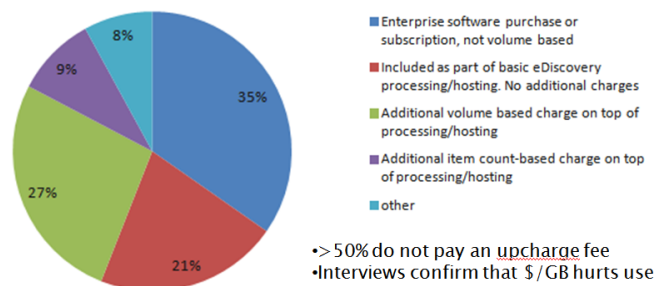
Law Firm Buyers – Biglaw litigation support is under pressure to process/analyze dramatically larger, diverse ESI collections under tightening litigation and regulatory deadlines. Interviews revealed strong resistance from senior partners to any machine learning review technologies that threaten the traditional high margin associate document review model. These conflicting pressures have focused RFP’s on eDiscovery processing platforms augmented by clustering analytics to enable a few techs to assess, process and organize incoming collections/productions for prioritized review.

Service Provider Buyers – Consolidation among the national service providers has driven down pricing for low margin ESI processing and hosting. Smaller regional providers have been forced to differentiate with concierge services or managed service dedicated contracts. You could say that the Alternative Fee Arrangement (AFA) trend has trickled down from the law firms into the eDiscovery market. Providers have to offer PC/TAR review to be competitive, even if it is only used on a tiny portion of matters.

Consumption Models

Surprisingly, over 50% of combined survey respondents do not pay an upcharge for analytic functionality. This metric should be tempered with the understanding that a large portion of respondents defined analytics as deduplication, email threading and similar processing optimization functionality. Thus ‘analytics’ can be seen as a normal part of basic ESI processing and included in the lower \$/GB fees.

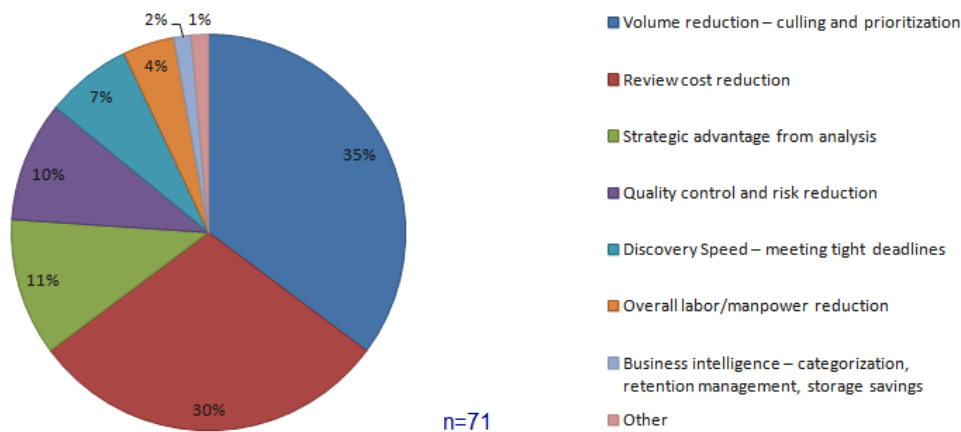
Analytic Consumption Models



Survey Metrics – A Limited View

The seven survey questions sought to understand the adoption rate, usage cases, perceived value, consumption models, direct users and the brand recognition of specific analytic offerings. The exact language and interpretation of the responses should be tempered by the fact that respondents had a wide range of definitions of ‘analytics’. This became readily apparent during interviews when discussing Question #2, “What portion of matters do you or your customers use some form of analytics on?” Most respondents included deduplication, email threading and other processing functionality in their responses. The interviews gave eDJ a chance to get specific about PC/TAR review rates (5-7% overall), a much more useful statistic when looking at the overall market adoption. The combined survey responses are found in Appendix 2 and the individual Consumer and Provider Survey Reports are available on the eDJ Group website to Participating Members.

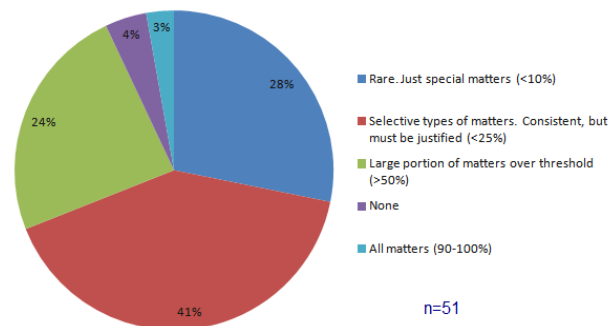
Question #1. What do you see as the primary selling point for using analytics in eDiscovery or Information Governance for you or your customers?



Providers overwhelmingly selected volume reduction, review cost and QA, while consumers also valued strategic advantage (ECA) and discovery speed. More importantly, interviews indicated that consumers perceived review cost savings from optimized/clustered review while providers interpreted review savings as coming from PC/TAR machine learning workflows.

2. What portion of matters do you or your customers use some form of analytics on?

As discussed, the responses include deduplication, email threading, facet navigation and other common functions, which limits the value of this question. eDJ was surprised at reports from providers that some clients do not deduplicate or even organize documents prior to traditional linear review.

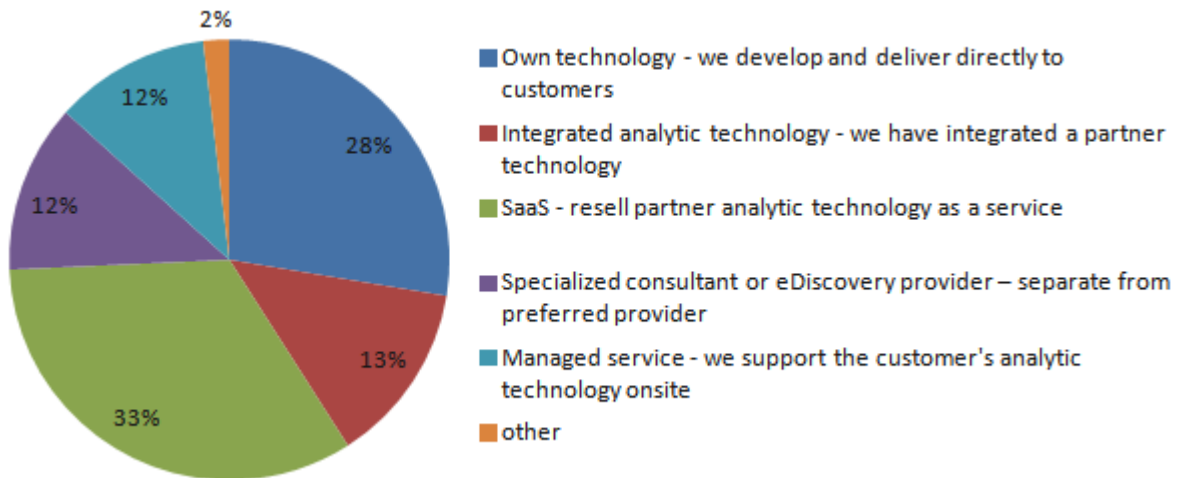


3. What general usage cases do your customers use analytics on?

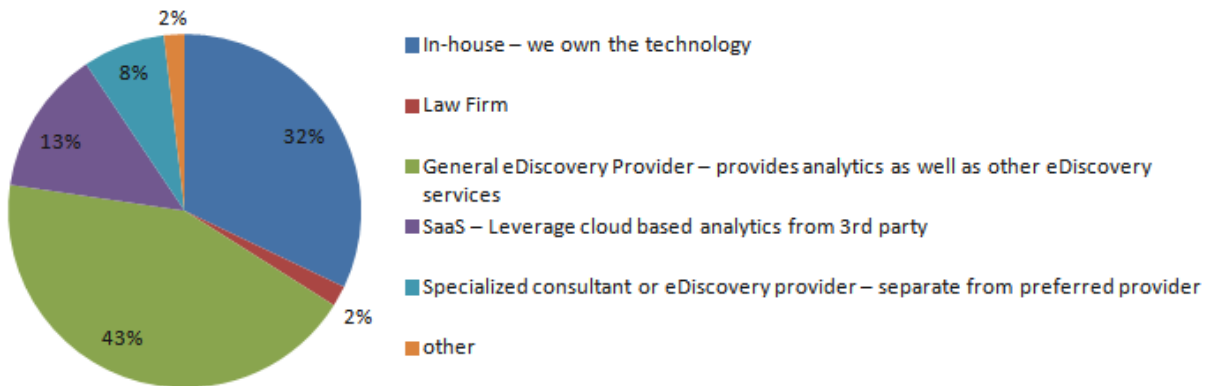
The dominant analytic usage cases are eDiscovery ECA, profiling and analysis of opposing productions. IG usage cases are much less common and interviews confirmed that most represent exploratory projects rather than full enterprise initiatives with board level backing. See detailed breakdown of combined responses in Appendix 2.

4. Who provides the analytics that you or your customers use?

Combined Responses:



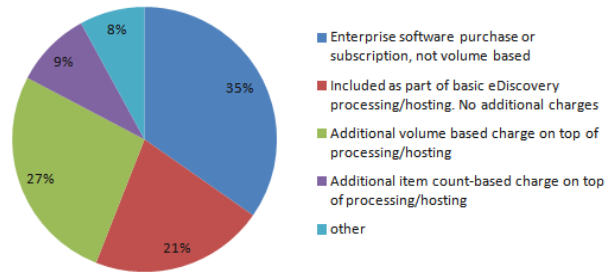
Consumer Only Responses:



eDJ was surprised that roughly 40% of combined responses indicated analytic software running behind the firewall. SaaS offerings seem to be on the rise, as demonstrated by Recommind’s conversion to a SaaS provider and Microsoft’s impending acquisition of Equivio. Overall, analytics continue to be consumed primarily from an external provider and eDJ does not expect that trend to change quickly.

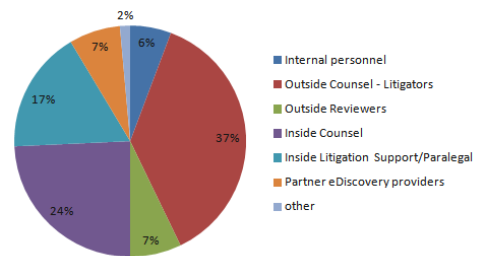
5. How do you purchase/consume your analytics?

Again, we see 56% of respondents consuming analytics without an additional upcharge on the basic system or service. Analytics are perceived as value-added functionality, but there is strong resistance to increased costs. eDJ has observed increasing downward pricing pressure on providers with OEM analytic license costs and expects many of those licensing agreements to convert to more flexible fixed cost models in response.



6. Who are the primary users/consumers of your analytic offerings?

Interview responses mildly contradicted the survey results. eDJ heard numerous reports that senior counsel (inside/outside) resisted direct use of the visualizations or software. Instead, they consumed the system results while relying on providers, lit support and junior legal staff to operate the actual technology. Counsel make the buying decisions and consume the output, but few seem to be comfortable navigating relatively complex search, filtering or review interfaces.



7. What analytic systems has your company deployed or tested?

This question was designed to measure brand awareness and market penetration by 40 offerings with analytic functionality. Respondents indicated if the product had been demonstrated, used in cases or otherwise managed for customers. Appendix 2 has the combined totals as an arbitrary brand awareness measurement, but the options to providers and consumers do not exactly align. eDJ recommends that readers desiring to really understand the metrics review the separate survey results or contact eDJ for a detailed review. The primary take away is that the top 5-7 providers dominate the analytics market in terms of brand awareness.

Consumer Interview Perspectives

Corporate Perspective

One surprising figure from the consumer survey was that almost 20% see the primary benefit as the strategic advantage from analysis. My interviews gave more depth to this answer with usage scenarios requiring firm and corporate lit support to triage early collections from key custodians so that counsel can make the risk-liability-cost decisions. This kind of early analysis has been at the heart of the value that ‘big data’ index providers who have been promising search and more from live data sources. Instead of performing ECA on the live systems, my respondents reported using targeted collections on

key document sets and some optimized searches. The important story is about consumers seeing value in resolving the case or the discovery as quickly as possible, rather than the traditional delaying tactics seen for too many years. As discussed above, corporate discovery teams are highly aware of the rising volume and cost associated with preservation and discovery requests. Analytics, profiling and PC/TAR offer some potential relief, but corporations struggle to leverage them on enterprise data in place.

Key Interview Feedback:

- No real upstream IG usage. Preliminary exploration only of analytics on live enterprise data sources. eDJ heard reports of a few government agencies using vendors for upstream analytics.
- The perception of front loading analytics costs conflicts with the expectation that almost every case will settle before trial. How do buyers justify potential savings before they understand the relative relevance richness, true scope of potential collections and quantify liability.
- Corporate interviewees believed that the impact from regulatory requests will exceed discovery costs in the near future. eDJ's corporate interviews focused on global corporations instead of smaller, private businesses, but provider interviews confirmed the trend.
- Regulators accelerating request deadlines and not accepting burden arguments. *"If you cannot review the email and files in time, we will take it all and use ProductX to find what we want."*
- Lit Support seeks to control volume prior export with standardized processing, exclusion filters and aggressive search terms. Raises concerns as to whether over-burdened Lit Support have the bandwidth to process massive preservation collections within a formal workflow that includes tool validation, formal workflow documentation and effective QA/QC.
- Counsel not interested in direct access to PC/TAR interface, but want LitSupport to run it without investing in support expertise.
- Simpler interface the better for complex machine learning technology.
- Will not purchase analytics separately or as mark up.
- 80% of analytic usage in investigation/regulatory.

Law Firm Perspectives

eDJ heard conflicting stories from law firm Lit Support and counsel on PC/TAR systems. Those in the discovery trenches want any tool to handle the flood of ESI from clients and opposing counsel. Partners and senior litigation counsel resist innovation because it threatens the traditional associate review revenue model and because they have legitimate risk concerns. The firms are the battleground where PC/TAR adoption is being fought. eDJ believes that most of the resistance will collapse when providers abandon volume based up-charges for analytics.

Key law firm feedback:

- Primary motivation for PC/TAR usage is speed. Many cited tightening regulatory deadlines.
- Firms need analytics early in the process, but find it hard to justify additional cost.
- Cited confusion and disbelief over PC/TAR marketing claims. Want a simple one-page ROI of linear review vs. accelerated review savings to help justify the cost to clients.
- Regulators still want it all if they can get it. Agencies invested in analytic tools to tackle big data.

- Multiple reports that PC/TAR with 95-99% target rates can result in production of 3-500% non-responsive docs. Plaintiffs demanding seed sets. Clients reject PC/TAR to protect confidential IP.
- Big focus on strategic advantage from ECA and early profiling.
- Mature corporate clients already cull during processing, so why pay firm for analytics?
- Co-counsel in large matters resisting PC/TAR due to loss of revenue and exclusion from training/review process.
- Even term “Predictive Coding” scares some counsel. Math terms such as recall, precisions and accuracy escapes majority of legal market.
- <20% of clients actually asking about PC/TAR. Even AmLaw 100 firms with on-premise PC/TAR are using it mainly for ECA and processing optimization with very few actual machine learning relevance reviews. They are using it to triage opposing productions.
- PC/TAR does not always save money. More about improving quality and speed. System and technology less important than the team running it. Money does not translate to quality.

Technology Agnostic Provider Perspective

Providers seem to be laser focused on optimizing review efficiency, while the consumer interviews reflected a much wider set of drivers for analytics. This makes sense when you consider the classic service provider role, but it means that some providers can miss the early strategic opportunities to resolve or minimize discovery. eDJ had multiple reports of short sighted providers and law firms resisting adoption of true PC/TAR for reviews for a wide variety of reasons. The use of analytics to optimize linear review sets enjoys far more widespread acceptance, though providers and law firms report having to make the ROI justification for the additional processing costs in most cases. This ‘Accelerated Review’ methodology was pioneered by Attenex and Stratify more than 10 years ago and it is much easier for counsel to understand how decisions are propagated in clusters. This transparency lowers adoption resistance and avoids some of the potential seed set disclosure traps reported by sharp eDiscovery litigators.

Key provider feedback:

- Reports that cross case designation/production/category history and context on frequent flier custodians dramatically improves review rates. Not strictly analytics functionality, but interesting finding from multiple interviews.
- The only 100% PC/TAR adoption story came from a managed client with a fixed cost contract that included analytics. All-in pricing with free consulting support translated to PC/TAR adoption. Isolated success story that was not found in other providers.
- General reports of low user maturity. Many PC/TAR and analytic functionality misconceptions. Expertise requirement scares off new users.
- Reports that the relatively high PC/TAR training effort (10%+, 10-20 review rounds) associated with some systems translates to adoption resistance and outright rejection despite documented savings. Just not worth the hassle and potential risk.
- Unconfirmed reports of SEC/DOJ fighting proposals for TAR protocols in certain regions.
- Example adoption metrics:

- 90% cases – email threading/near dup exclusion
- <1% actual review cases PC/seeding despite long term provider investment in PC technology/expertise.
- End clients make analytic decision/budget. Firms see analytics stealing associate review revenue.
- Analytics used more for investigations than actual production review.
- Use social network, timeline functionality heavily, but as part of manual workflow rather than automated processing.
- Volume pricing for processing and hosting keeps falling and makes it harder to add analytics up-charges.
- Increased interest in post- and plaintiff production analysis driven by low risk and effort. Does not require 99% recall. Only looking for key documents/evidence.
- Clients on marketing information overload. Some providers retreating from PC/TAR evangelism based on marketing fatigue and desensitization. Stopped sales go-to-market in the face of customer indifference.

eDJ Perspective

PC/TAR Adoption Rate

Between the Analytic Adoption surveys, product briefings and extended interviews with cutting edge LitSupport, attorneys and providers actively using analytics, the vast majority of consumers and cases are not yet ready for to use PC/TAR to make relevance decisions on productions. This reality conflicts with everything seen at Legal Tech and other eDiscovery marketing channels. eDJ does not question the potential effectiveness of the technology or the expertise of the subject matter experts that seem to be required to manage these machine learning reviews; whether active, passive or just plain black box driven. The initial survey metrics on analytics used in review seemed quite encouraging⁸ until the live interviews. The combined survey questions asked, “What portion of matters to you or your clients use some form of analytics on?” While you could interpret the combined responses to to validate the widespread use of PC/TAR, the follow up interviews made it clear that the market has a much broader definition of ‘Analytics’ than the marketers pushing PC/TAR. Interviews selected respondents with deep eDiscovery experience that had used PC/TAR on actual cases. Every one of them quickly differentiated between the use of analytics to prioritize or optimize collections prior to review versus the actual use of any kind of machine learning PC/TAR technology. The use of ‘Accelerated Review’ during or just post processing to prioritize, cluster or otherwise optimize the collection is now well accepted and seems to be how the vast majority of survey respondents interpreted Question #2.

When asked about the actual use of PC/TAR for real machine learning review, the numbers plummeted. eDJ’s best estimation is that only 5-7% of matters that reach the review stage (remember that most cases settle well before review) actually use some form of PC/TAR. And that level of adoptions seems to

⁸ Question #2 in Appendix 2

be driven primarily by a small number of critical matters under tight deadlines or by a few very large defendants who have negotiated ‘all in one’ preferred provider agreements that wrap analytics into some kind of fixed fee/rate arrangement. Are there companies that use PC/TAR on every review? eDJ found these exceptions, but these were the earliest of the adopters and do not represent the broader market. Provider interviews all concurred that we are still in an educational sales cycle. Many providers have backed off from actively selling or promoting PC/TAR in the face of resistance from senior law firm partners and confused corporate customers.

PC/TAR Adoption Challenges

PC/TAR has dominated eDJ’s recent briefing sessions with providers and consumers alike. Consumers want to clarify the terminology, technology and market hype surrounding recent cases. Providers have expressed frustration with the portrayal of PC/TAR as some kind of ‘Easy Button’ that will magically reduce review expense by 95%. eDJ heard second hand stories like, “*But VendorX says his system only needs to train with 5%.*” Early PC/TAR innovators like DiscoverReady’s CEO Jim Wagner long ago understood that, “*It’s not the technology. It’s the people and process.*” That can be a complicated message for a relatively unsophisticated consumer who reads sponsored blogs instead of actual transcripts. Discovery and review cease to be easy or routine as the volume and composition of potential collections exceed the ability of a single reviewer to manually code every item. Beyond simple relevance the additional complexities of privilege in TAR keep coming up in our briefings.

The comingled business and legal roles of many corporate counsel and experts challenges the old assumption that including an attorney on communications automatically conveys privilege protection. This is old news and counsel has become familiar with 2nd and 3rd pass reviews to adjudicate these complicated issues. But how do PC/TAR technologies and workflows deal with embedded privilege in email threads, attorney comments in spread sheets or secondary work product reports? Most clustering, propagated or predictive review systems are optimized to train for one ‘category’ at a time, that being relevance. Training for individual issues or privilege can be done with separate seed sets of known documents and iterative sample sets, but the stakes for missing a single privileged document buried in the final production set can be serious, just ask Google⁹.

Some challenges with PC/TAR:

- Custodial based preservation results in ever broader collections.
- Broad collections of unrelated ESI creates very low recall and relevancy rates.
- Most PC/TAR systems have a minimal threshold of relevant documents to be effectively trained.
- Clustering, categorization and other linguistic grouping technologies can be overwhelmed by non-relevant similarities in raw collections. (*they can’t see the trees for the forest*)
- The comingling of business and legal roles in corporate communications mean that potential privilege contaminates a large portion of ‘decision’ documents. (*the same ones that TAR is being trained for*)
- PC/TAR is just one tool to leverage within a structured discovery process.

⁹ <http://www.law.com/jsp/lawtechnologynews/PubArticleLTN.jsp?id=1202542577684&slreturn=1>

Every eDJ conversation with PC/TAR experts, providers and consumers reinforces the need for adaptive, integrated solutions that combine search, sampling, profiling and PC/TAR for optimized, defensible results.

The market is over-saturated with promises like *“Review only 5,000 items!”*¹⁰ and *“Save 75% of review cost!”*¹¹ Although every corporate interviewee had used some kind of machine learning in a few cases, none felt that PC/TAR was appropriate for every matter. The dominant reason for using PC/TAR was to meet impossible deadlines or overwhelming volumes of raw ESI. PC/TAR was not considered easier or cheaper when you factored in the up-front cost of the analytic processing (\$130-200/GB extra), technical experts to manage the process and the cost to negotiate the protocol with the opposing party. So why did they go through all that trouble if they did not anticipate real savings? Because they did not have a choice. Many cited very tight time frames in recent regulatory requests. Others talked about being ‘dump trucked’ by huge opposing productions just weeks before depositions. Law firm interviews revealed a struggle between conservative partners and progressive litigation support evangelists. The eDiscovery providers gave explicit descriptions of conversations with law firm partners with quotes like, *“Your ROI for this stuff comes out of my pocket!”*

eDJ’s Top Ten Reasons Why NOT PC/TAR:

1. Perception that PC/TAR costs front load the discovery cost for matters that will settle before trial.
2. High resistance to analytic up-charges. Have to justify them on every matter, so go with path of least resistance.
3. Complexity of systems and fear that counsel will not be able to defend what they do not understand.
4. Customers on information overload. Marketing fatigue and growing customer indifference.
5. Perception that PC/TAR reinforces known relevant selection and misses unknown/new documents.
6. Rumors of SEC/DOJ in some areas fighting PC/TAR proposals.
7. Realization that 95-99% recall in PC/TAR training will result in 300-500% production size. Exposure of large volumes of non-relevant ESI a serious concern for companies facing serial plaintiffs on ESI fishing expeditions.
8. Mature corporate customers already cull and optimize during collection or processing. If they can achieve substantial savings prioritizing/clustering review sets, why pay for actual PC/TAR analytics?
9. Counsel does not want to operate PC/TAR systems. Wants Litsupport or provider to run it.
10. PC/TAR takes money from the firm. Takes away associate jobs.

Are most of these market perceptions true? While eDJ does not believe that they are true, they all contribute to the slow adoption rates for actual use of machine learning technology in traditional discovery review for production. Since this phase of the eDiscovery lifecycle is so frequently cited as being responsible for 75% of the cost of discovery, you can see why many providers had become disenchanted with sales evangelism focused on PC/TAR. Providers know that they must be able to host and support some kind of PC/TAR, but with kCura’s Relativity becoming the default review platform for

¹⁰ <http://orcatec.com/2013/01/22/global-aerospace-predictive-coding-results-approved-by-judge-for-1st-time/>

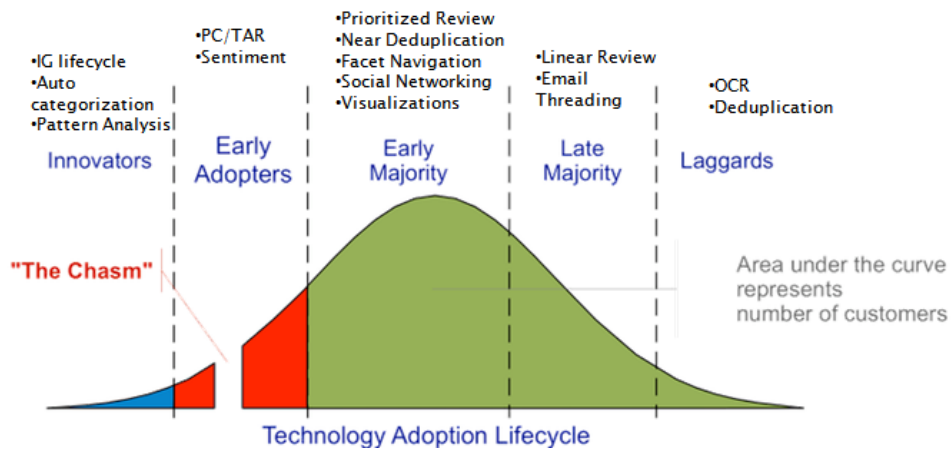
¹¹ <http://tanenholzlaw.com/predictive-coding-cost-savings>

providers that becomes easy. They have backed off pushing PC/TAR in the face of direct negative feedback from some firms and resistance from corporations who expect the matters to settle. The use of these technologies for ECA, investigations, regulatory responses and analysis of opposing productions will continue to slowly eat away the fear, uncertainty and doubt that have kept the adoption rate for machine learning review at such a low level.

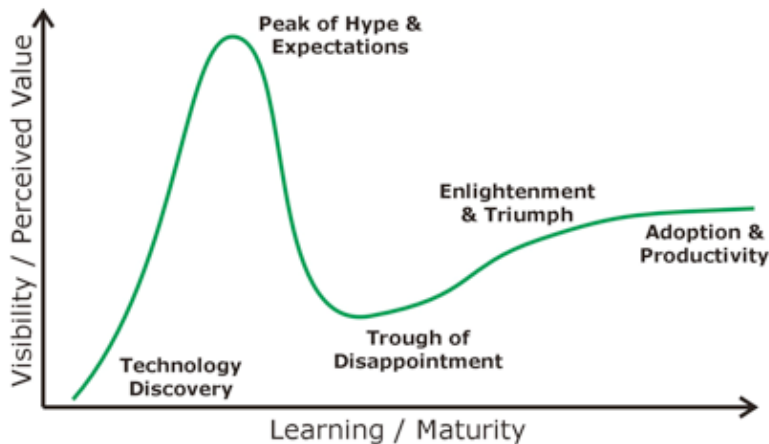
The Sweet Spot of the Maturity Curve

A comparison of Geoffrey Moore’s technology adoption life cycle model¹² and Christopher Rollyson’s Adoption Curve for Web 2.0¹³ can shed a bit of light on the ‘Trough of Disappointment’ that eDJ’s research uncovered.

Analytic Maturity Spectrum



ADOPTION CURVE



¹² Geoffrey Moore's Crossing the Chasm technology adoption life cycle model - http://en.wikipedia.org/wiki/Crossing_the_Chasm

¹³ <http://socialmedia.biz/2009/04/18/web-20-adoption-curve-2009-2015/>

Analytic buyers should understand their own tolerance for risk and their ability to leverage these technologies before investing in eDiscovery or IG analytics. Although cloud SaaS offerings have lowered implementation cost and effort, buyers should never underestimate the difficulty of migrating critical data assets or active cases if the chosen platform does not live up to the sales hype. So what is the market's 'sweet spot' where buyers are getting the most value?

Consumers and providers agreed that the application of analytics to assess, process and prioritize ESI for review has the least resistance and the clearest ROI at present. The clearest success stories involved preferred provider relationships where the analytics technologies were included in the base cost of discovery. For some reason, counsel seems to have all kinds of issues with an automated system that identifies and excludes 'non-relevant' categories of data on the basis of machine learning, yet they do not seem to care when processing rules filter out lunch notices, Face Book forwards and other categories. Although this could be nothing more than willful ignorance or "*don't ask, don't tell*" behavior, eDJ has seen no significant case law, articles or market noise about pre-review filtering/processing. Interviews, marketing case studies and client metrics show consistent 300-500% increases in review performance from a wide variety of optimization clustering, flagging and recommendation techniques. This allows retained counsel to demonstrate savings and efficiency without eliminating the large scale review process.

Analytic Market

eDJ believes that add-on or embedded analytic functionality is required for any technology that supports the central eDiscovery lifecycle on premise, hosted or in a true cloud SaaS environment. Simple Boolean indexed search is no longer sufficient to meet consumer demands and tackle increasingly large, diverse ESI collections. Processing and review platforms must enable a relatively small number of primary users to assess, filter, group and effectively manage these collections. Clustering, threading and facet extraction functionality cannot be developed or integrated overnight, so eDJ believes that partnered analytic engines will continue to dominate the 2015 market. Although many review platforms have incorporated open source or home grown clustering and visualizations, few can match the focused development investments of the pure analytic engines. kCura's Relativity is the best example of how an eDiscovery platform can partner with a wide range of analytic engines to meet customer demands.

What about the IG market for analytics? Despite a couple early forays into the IG market and the recent acquisition offer by Microsoft for Equivio, eDJ sees only very early adopters with strong regulatory or business intelligence mandates making investments in IG analytics. The 'Big Data' hypecycle is well under way, but providers report long educational sales cycles that rarely receive executive backing for more than pilot programs. By itself, "Information Governance" is an undefined data lifecycle ideal without concrete customer pain points and value propositions. eDJ expects that successful providers will target very specific analytic initiatives such as smart migrations, defensible deletion, PST elimination, data maps, in place legal holds, compliance monitoring and security enhancements in 2015.

Executive Summary

The term ‘analytics’ covers a wide variety of technologies applied to usage cases across the Information Governance and eDiscovery lifecycles. Consumers with specific pain points have embraced analytic service offerings to overcome increasingly large and diverse data repositories, while hesitating over on-premise adoption. eDJ believes that core clustering, profiling and visualization functionality is essential for every IG or eDiscovery platform covering a broader part of the lifecycle. The key factors to



successful adoption and utilization of analytics include a concrete goal for the functionality, plain ROI value proposition, internal/external expertise and documented workflow to reduce risk. Inquiries from eDJ consulting clients initiated this research and are typical of many enterprise teams wondering if analytics can reduce their rising costs, legacy repository volumes and provide strategic insight for early resolution of risk matters. Consumers want an analytic ‘Easy Button’, but distrust market hype.

Predictive Coding/Technology Assisted Review (PC/TAR) has dominated the eDiscovery hype cycle for the last two years. Evangelist providers with well funded marketing machines have cranked out webinars, articles, case studies, conference panels and advertisements proclaiming the end of traditional linear review. Senior firm partners and inside counsel have deep, different concerns about the using machine learning to make critical review decisions without some kind of eyes on every item produced. eDJ estimates that only 5-7% of matters reviewed are using some variation of machine learning in the actual review process. In contrast, eDJ estimates that over 90% of matters use analytics in processing or pre-review organization to increase review rates (3-500% reported) and quality. This ‘review acceleration’ was typically included in normal processing and project management costs, whereas PC/TAR services can require approval of additional up front charges on matters that almost always settle before trial. In addition, conservative 95-99% accuracy thresholds usually result in large overproduction of non-relevant documents and open the door for demands for production of all non-privileged training documents (which can be 15-35% of the collection). The eDJ bottom line is that PC/TAR adoption still faces significant resistance and legitimate concerns, though eDJ does believe that adoption will continue to grow as provider solutions, judiciary, regulators and parties continue to mature.

Appendix 1: eDJ Matrix Analytic Offerings

- AD eDiscovery
- Autonomy IDOL
- Axcelerate eDiscovery
- Brainspace (formerly PureDiscovery)
- Case Logistix
- Catalyst
- Catelas
- Cicayda
- Clearwell
- Content Analyst
- Daegis Edge
- Digital Warroom
- DigitalReef
- Discovery360 DataMapper
- Equivio
- Exterro Fusion
- Hadapt
- Intella Pro
- Kroll
- Lexalytics
- NexLP
- NuixOmniX
- OrcaTec
- Palantir
- Percognate
- PlanetData Analytics
- Prolorem
- RDC Analytics
- Recommind
- Relativity
- Ringtail
- Servient
- StoredIQ
- TunnelVision
- Viewpoint
- Xera
- Zylab

Example functional comparison of 8 top analytics offerings from www.eDJGroupInc.com. Features may be updated or change.

eDJ Matrix Comparison Chart

Show by Category: Analytics Compare by: Offering

[back to results](#) [save](#)
[download report](#)

eDJ REVIEWED	✓	•	•	•	•	•	•	•
Identify-Data Map	•	✓	✓	•	•	•	•	•
Analysis-ESI Sampling	✓	✓	✓	✓	✓	✓	✓	✓
Process-Inventory	✓	✓	✓	✓	✓	✓	✓	•
Process-Autocode-Objective	•	✓	✓	•	✓	•	•	✓
Process-Deduplication-Hash	✓	✓	✓	✓	✓	✓	•	•
Process-Communication De-Duplication	✓	✓	✓	✓	✓	✓	•	✓
Search-Concept	✓	✓	✓	•	✓	•	✓	✓
Process-Near Duplicate ESI	✓	✓	✓	✓	✓	✓	•	✓
Process-Deduplication-Email Thread	✓	✓	✓	✓	✓	✓	✓	✓
Analysis-Linguistic	✓	✓	✓	•	✓	•	•	✓
Analysis-Domain	✓	✓	✓	✓	✓	✓	✓	✓
Analysis-Social Networking	✓	✓	✓	✓	•	✓	•	✓
Analysis-Chronological	✓	✓	✓	✓	•	✓	•	✓
Review-TAR Conceptual	✓	✓	✓	✓	✓	•	✓	•
Analysis-Skin Tone	•	•	✓	•	•	✓	•	•
Analysis-Performance Metrics	•	✓	✓	•	✓	•	✓	•
Review-TAR Predictive	✓	✓	✓	✓	✓	•	✓	•
Analysis-Near Duplicate-OCR	✓	✓	✓	✓	✓	•	✓	✓
Search-Facet Navigation	✓	✓	✓	✓	✓	✓	✓	✓
Process-Foreign Language Identification	✓	✓	✓	✓	✓	✓	•	✓
Analysis-Custodian Identity	✓	✓	•	✓	•	•	✓	•
Review-TAR-Clustered	✓	✓	✓	✓	✓	✓	✓	•
General-Analysis	✓	✓	✓	✓	✓	✓	✓	✓
Analysis-Visualization	✓	•	✓	✓	✓	✓	•	✓
Analysis-GPS Mapping	•	•	•	•	•	✓	•	•
IG-Classification	•	•	✓	•	✓	•	•	✓
Analysis-Image Categorization	✓	•	•	•	•	•	•	•
Analysis-Dashboard	✓	•	✓	✓	✓	✓	✓	•
Analysis-Conceptual	•	•	✓	✓	•	✓	✓	✓
Analysis-Similarity	✓	•	✓	✓	•	✓	✓	✓
Analysis-Cluster-Soft	•	•	•	✓	•	•	✓	✓
Analysis-Context	•	•	•	✓	•	•	✓	✓
Analysis-Event Map	✓	•	•	•	•	•	•	•
Process-Autocode-Subjective	✓	•	•	•	•	✓	•	✓
Analysis-Machine Learning	✓	•	✓	✓	•	✓	•	✓
Analysis-Profile-ESI	✓	•	•	✓	•	✓	•	•
Analysis-Sentiment	•	•	✓	•	•	•	•	•

Appendix 2: Aggregate Survey Responses

1. What do you see as the primary selling point for using analytics in eDiscovery or Information Governance for your customers?

Value	Count	Percent
Volume reduction – culling and prioritization	25	35.2%
Review cost reduction	21	29.6%
Strategic advantage from analysis	8	11.3%
Quality control and risk reduction	7	9.9%
Discovery Speed – meeting tight deadlines	5	7.0%
Overall labor/manpower reduction	3	4.2%
Business intelligence – categorization, retention management, storage savings	1	1.4%
Other	1	1.4%

2. What portion of matters do your customers use some form of analytics on?

Value	Count	Percent
Rare. Just special matters (<10%)	20	28.17%
Selective types of matters. Consistent, but must be justified (<25%)	29	40.85%
Large portion of matters over threshold (>50%)	17	23.94%
None	3	4.23%
All matters (90-100%)	2	2.82%

3. What general usage cases do your customers use analytics on?

Usage Case	Combined	Percent
IG – Enterprise business intelligence and categorization	7	1.60%
IG- Compliance monitoring	7	1.60%
IG- Audit/risk analysis	2	0.46%
IG-Operations optimization	1	0.23%
IG-Retention enablement	8	1.83%
IG-Security/Investigations	13	2.97%
Discovery - ECA/Identification/Scoping/Negotiation Support	38	8.70%
Discovery - Selective Preservation	7	1.60%
Discovery - Collection scoping	13	2.97%
Discovery - Processing/Filtering	38	8.70%
Discovery - Relevance and Search Criteria	50	11.44%
Discovery - Exclusion Criteria	38	8.70%
Discovery – Review - Clustering/Grouping (Accelerated Review)	48	10.98%
Discovery – Review - Machine Learning/TAR/Decision Propagation	42	9.61%
Discovery – Review - Predictive Coding/Recommendation	44	10.07%
Discovery – Review - Quality Control	41	9.38%
Discovery – Review - Opposing production analysis	39	8.92%
other	1	0.23%

4. Who provides the analytics that your customers use?

Value	Combined	Percent
Own technology - we develop and deliver directly to customers	29	27.6%
Integrated analytic technology - we have integrated a partner technology	14	13.3%
SaaS - resell partner analytic technology as a service	35	33.3%
Specialized consultant or eDiscovery provider – separate from preferred provider	13	12.4%
Managed service - we support the customer's analytic technology onsite	12	11.4%
other	2	1.9%

5. How do you purchase/consume your analytics?

Value	Combined	Percent
Enterprise software purchase or subscription, not volume based	26	37.1%
Included as part of basic eDiscovery processing/hosting. No additional charges	16	22.9%
Additional volume based charge on top of processing/hosting	20	28.6%
Additional item count-based charge on top of processing/hosting	7	10.0%
other	6	8.6%
Provided for free by law firm	0	0.0%

6. Who are the primary users/consumers of your analytic offerings?

Value	Count	Percent
Internal personnel	4	5.7%
Outside Counsel - Litigators	26	37.1%
Outside Reviewers	5	7.1%
Inside Counsel	17	24.3%
Inside Litigation Support/Paralegal	12	17.1%
Partner eDiscovery providers	5	7.1%
other	1	1.4%
Compliance/Security or other business users	0	0.0%

7. What analytic systems has your company deployed or tested?

Brand Awareness	Demo/Use	Percent	Case Use
Relativity	48	14.95%	41
Equivio	39	12.15%	20
Clearwell	34	10.59%	17
Nuix	34	10.59%	14
Recommind	29	9.03%	10
Content Analyst (CAAT)	23	7.17%	12
Catalyst	19	5.92%	9
Viewpoint	19	5.92%	9
OrcaTec	17	5.30%	5
Kroll	16	4.98%	7
Ringtail	16	4.98%	7
Axcelerate eDiscovery	15	4.67%	7
Exterro Fusion	14	4.36%	1
Autonomy IDOL	13	4.05%	2
Xera	12	3.74%	5
Case Logistix	10	3.12%	2
Brainspace (formerly PureDiscovery)	10	3.12%	5
TunnelVision	9	2.80%	3
AD eDiscovery	7	2.18%	2
Digital Warroom	7	2.18%	1
Zylab	6	1.87%	1
Daegis Edge	6	1.87%	0
Cicayda	6	1.87%	0
Servient	5	1.56%	3
Discovery360 DataMapper	4	1.25%	1
StoredIQ	4	1.25%	1
DigitalReef	3	0.93%	0
Palantir	3	0.93%	0
NexLP	3	0.93%	0
OmniX	3	0.93%	0
Liquid Litigation Management, Inc.	3	0.93%	1
RDC Analytics	2	0.62%	1
Catelas	2	0.62%	0
PlanetData Analytics	1	0.31%	0
eZVUE (ECA tool)/eZReview ADP (TAR tool)	1	0.31%	0
IPro	1	0.31%	1
Hadapt	0	0.00%	0
Intella Pro	0	0.00%	0
Lexalytics	0	0.00%	0
Percognate	0	0.00%	0
Prolorem	0	0.00%	0

For detailed breakdown of demo and actual usage see original survey data.

Appendix 3: New eDJ Matrix Features and Descriptions

Process-Autocode-Subjective

Extraction of key sentences and data elements to construct a summary description of the document. Auto summarizing

Analysis-Cluster-Concept

Ability to extract and present conceptual clustering for visualization, navigation, search or TAR review.

Analysis-Cluster-Similarity

Ability to identify, group and retrieve items based on similarity or near-duplication. Can be used find similar items, cluster, deduplicate or condense iterative conversation strings.

Analysis-Cluster-Soft

Items can be clustered into multiple groups instead of just one category. Also called overlapping clustering.

Analysis-Context

Ability to cluster or search based on contextual elements of ESI such as time, location, participants and more.

Analysis-Event Map

Ability to reconstruct diverse facts, ESI and decisions to present an event from different perspectives in time, location and data systems.

Analysis-Machine Learning

Ability of the system to learn from user designations to create relevance criteria or be incorporated into accelerated TAR review methods. This feature can be considered a broader functionality than the Review-TAR features with upstream usage cases.

Analysis-Profile-ESI

Generate profile report or facet navigation of ESI collection that gives statistical breakdown of file types, dates, owners, associated custodians, locations and other meta data facets. The profile can be static or support dynamic navigation and actions.