

9 Top eDiscovery Trends for 2012

by Christine Taylor

So many people come out with end of the year “Trends” articles that I sometimes skip it. However, there is so much rapid morphing in the eDiscovery industry that if I didn’t share my 2012 trends I’d be a slacker. We can’t have that.

The following 9 points are some of the trends you need to know about the complex field of eDiscovery coming into 2012.

1. Social Media

Ah, social media... potentially ruinous ESI run amok. When people think about social media they mostly mean Facebook and Twitter with an occasional nod to LinkedIn. With millions of people Face-booking, Tweeting and Linking away, information related to disputes and investigations is bound to end up online. In addition to the Big Three, specialized social media sites are proliferating and aggregators multiply large volumes of shared data.

The courts are not blind; they are aware that attempting to meaningfully search and collect data in social media is a huge challenge. They are generally sympathetic to proportionality arguments but not to the point that they allow litigants to simply skip social media searches. But traditional collection practices don’t cut it. Forget trying to image Twitter servers or getting Facebook to run collections on their data storage. Even Facebook’s self-collection tool is extremely limited and can only be run from a user’s own login account. Nor do third-party screen captures meet defensibility standards.

There are some social media tools that work in limited areas. Some collect using Twitter, Facebook and/or LinkedIn’s metadata schemes (all of which are incompatible with each other) but their scope is limited. Other software products are set to regulate user-created data for compliance; useful but not data collection. This field is wide open to technical innovation and eDiscovery software developers know it.

Our take: Product development will be hot and heavy as eDiscovery vendors seek to develop products that will collect and preserve relevant data from social media sites. In the meantime, we strongly suggest that large matters require litigants to spend the money on social media collection experts. Their mix of proprietary technology and consulting expertise is the way to go for now while vendors develop more sophisticated in-house methods for social media searches.

Vendors we are following: [XI Discovery](#), [FTI](#), [Smarsh](#)

2. Managing the Corporate e-Discovery Workflow

Wanting to manage the eDiscovery workflow is not new. The ability to do it well in-house (i.e. without using expensive consultants) is improving somewhat. But because the eDiscovery workflow is already large and complex, progress can be slow.

Workflow remains broadly categorized by search and collection, preservation, early analysis and processing on the corporate side; followed by review, production and presentation on the law firm side. But growing data and a dearth of in-house eDiscovery management tools are making supervising the process difficult. Today there are three general approaches to managing eDiscovery workflow -- as opposed to not managing it and hoping for the best.

- The first approach is to hire law firms or specialized consultancies to manage workflow. For large matters with high stakes, this is the right way to lower the risk of poor collection, preservation, timeliness and review.
- The second method is using technology tools that manage workflow in-house. These usually depend on operating with a single-vendor platform, since the more disparate the eDiscovery tools the less well the management tool will work.
- The third method of workflow management is based on internal eDiscovery best practices and methodologies instead of depending on outside consultants, law firms or vendors. This requires a tightly integrated workflow and project management process that operates across multiple departments. It allows corporations to leverage existing eDiscovery purchases.

Our take: Organizations must ask themselves what resources they need to manage eDiscovery workflow. Simple and infrequent matters can be handled internally with review passed on to the law firm. Workflow management can be done using an eDiscovery project manager and/or single-platform eDiscovery management software such as AccessData or Clearwell. Large and frequent matters are best handled by outside professionals from a consultancy or law firm. In all cases we strongly suggest that corporations develop internal eDiscovery best practices using cross-departmental teams. Just don't leave eDiscovery workflow management to chance; this never turns out well.

Vendors we are following: [Integreon](#), [AccessData](#), [Symantec Clearwell](#)

3. More Multi-Matter Management

Multi-matter technology is not new but interest is growing on the corporate side. Multi-matter management refers to preserving attorney work product and privilege and applying it to additional matters. The practice cuts down dramatically on the cost and length of a review. Corporations want this for obvious reasons but it is not exactly on the top of the law firms' Greatest Hits lists. Re-using intellectual capital cuts down severely on the amount they can charge for doing a new review for each matter, yet corporations are getting quite insistent on using multi-matter products to cut costs.

And they're right. Savings occur by leveraging previous privilege decisions and by tagging and commenting on a similar matter's responsive documents. And multi-matter re-use also helps to support quality control since no two reviewers will make the exact same decision across an entire review set.

Our take: Law firms will have to change their pricing structures to accommodate work product re-use. Of course we recognize that law firms must preserve profit. Law firms can immediately redress some lost revenue because there is not as much need to hire corps of reviewers for

similar matters. They can also make a good deal of revenue on newer or more unique matters. And of course law firms can always make a selling point out of multi-matter management usage.

Vendors we are following: [CaseCentral](#), [Daegis](#)

4. eDiscovery in the Cloud

On the face of it, the cloud is just another data storage target and application-delivery mechanism. However, there are serious legal and regulatory implications for some approaches to cloud-based storage. And this is exactly where the rosy predictions about cloud storage fail: security, privacy and defensibility can be difficult issues in the cloud.

However, not all cloud-based eDiscovery is created since not all of it represents big challenges. eDiscovery in the cloud falls into three distinct camps: 1) using the cloud for eDiscovery application delivery (SaaS), 2) archiving data to the cloud and contracting with the hosting provider to run eDiscovery processes, or 3) companies who store active data in the cloud that are subject to security, privacy and other regulatory actions. The first activity is low risk since data stays behind the corporate firewall. Software glitches happen and can impact productivity – for example Google Apps failures – but in the long-term customer impact is minimal. The second activity is also low-risk if the customer has done their homework and chosen an eDiscovery service provider with a secure data center and excellent track record. They are not storing primary data but archived data, itself single-instance copies of backed up data. The third activity however is riskier by far than the first two. Even where the cloud provider is trusted, such as Google or Amazon, service level guarantees for the enterprise are notoriously poor. And these services also have few mechanisms in place to report on physical data locations to their customers, which can be a serious defensibility issue.

Our take: Cloud-based application delivery and archival hosting with eDiscovery represent excellent returns on investment. However, multi-national corporations need to practice excellent data governance when data centers are located in different countries with differing privacy and eDiscovery laws. And organizations that are storing primary data in the cloud should very carefully consider security, longevity and service agreements when researching cloud storage vendors.

Vendors we are following: [CaseCentral](#), [HP Autonomy](#), [Orange LT](#), [Google](#), [Amazon](#)

5. Big Data Collection

Talk about data growth: projections see the world's digital data doubling every 1 to 2 years. It already averages an increase of 1-1/2 times a year. These massive growth rates build geometrically, resulting in nearly inconceivable amounts of data with no end in sight.

This leads to huge questions and issues around how massive data impacts hardware and software, data oversight, and of course eDiscovery and GRC. On the collection front big data particularly presents big challenges. The sheer size of data volumes slows down collection efforts, which are also impacted by different applications and storage locations. For example, email is well covered by existing eDiscovery tools and SharePoint and files are moderately well

covered. But cloud-based tools are sadly elementary and the explosively growing social media eDiscovery is in its infancy. Some vendors are looking to the intersection of business intelligence (BI) with information governance in order to provide automated analytics, reporting, responsive data movement, project level workflow insight and more.

Our take: Data collection tools are improving but fast-growing data presents ongoing challenges around multiple storage repositories, cloud-based storage, social media and collaborative digital creation. Collections tools must develop and grow in synch with these trends, which is both a challenge and an opportunity for collection vendors. It's a challenge because the technological barriers to collection are serious. It's an opportunity because the need is there and it's growing fast.

Vendors we are following: [EMC](#), [Oracle](#), [StoredIQ](#), [Guidance](#), [AccessData](#)

6. Predictive Coding

Predictive coding, which uses sampling to build and perform data set analysis, is quite a good tool for accelerating some manual analysis and review tasks. It is also not new: several vendors use predictive coding to accomplish eDiscovery tasks like analysis clustering, prioritization, threading or categorization. These activities shorten the review cycle by cutting down on unresponsive documents and organizing results into logical screens for reviewers.

In contrast, predictive coding for automated review seeks to replace manual review efforts for time and cost savings and for improved accuracy. This implementation of predictive coding learns progressively from expert attorney review examples, applies that learning to large review sets, and returns statistical samples for quality control. Ideally it will apply defensible review to large data sets in a fraction of the time that a team of reviewers could do it. In practice it is not nearly this easy, but it certainly holds promise for several review tasks.

Our take: Predictive coding for review is very promising but there is no case law for machine-automated document review. This is a serious problem for companies pushing predictive coding as review automation. We see three major drivers that will push for broader adoption: 1) established precedence, which will take some highly motivated trailblazers to accomplish; 2) corporate clients demanding a radical reduction in review costs and time, and 3) law firms building a competitive review practice who are willing to experiment with predictive coding.

Vendors we are following: [Recommind](#), [Catalyst](#)

7. Pressure towards eDiscovery Cost Reduction

eDiscovery has always been expensive and that will never change, but corporations should pursue cost-cutting measures anyway. As data grows, traditional processes just keep on getting more expensive and will necessitate new processes and evolving cost models.

Corporations look to two primary pricing models that also affect vendor bottom lines: total cost and line-item. The latter involves mapping a line-by-line process and expense chart and locating the lowest cost services required for each step, then estimating total cost. Invariably it's higher

than planned because the eDiscovery process is fluid and highly changeable, requiring many mid-course corrections between multiple vendors and service providers. Experienced project managers will build in a financial cushion to mitigate nasty financial surprises.

The total cost model is good work when you can get it. It depends on a single eDiscovery provider who can provide software, consulting, and workflow management for the specific stages required. The vendor may partner with outside entities but the corporate client sees the single price. (Law firm costs must be added on unless the law firm itself is offering the service.) The estimate may depend on documents, gigabytes and/or custodians involved. Total cost can be an economical method *if* the organization does not need to leverage existing eDiscovery products or alternate vendor relationships.

Our take: We believe in using a combined line-item and total cost approach. Line-item pricing benefits corporations with existing eDiscovery investments, relationships and eDiscovery workflow management expertise. Even these corporations can adopt total costing for later stages it makes no sense for them to do, such as processing-to-review-to-production. On the vendor side, eDiscovery product and service vendors are looking at ways to provide flexible pricing with caps so customers know what they're getting. Yet ultimately eDiscovery customers need to accept a simple fact of life: costs can be controlled but eDiscovery is not cheap and it never will be.

8. Continuing Consolidation

Some analysts are “predicting” widespread eDiscovery consolidation, to which I answer: no one needs a crystal ball to know that. It’s happening already. Over the next couple of years we will see the majority of consolidations as large storage and server companies eat up... I mean, thoughtfully acquire... smaller eDiscovery outfits.

This is not to say that these companies will necessarily do it *well*. I still don’t know what Iron Mountain was thinking when it released Mimosa and Stratify to Autonomy so soon, nor how HP intends to absorb Autonomy now. And Symantec and Clearwell? Their cultures may be friendly but their product lines and customers are very, very different. Having an API to Clearwell in Enterprise Vault does not a combined product line make. Also, with cloud and social media concerns growing – not to mention growing data of every type – the big guys will have to scramble to keep their acquired eDiscovery product sets current with marketplace needs.

Our take: Acquisitions and consolidation are a fact of life in a maturing market. I do not have a problem with storage and server companies acquiring eDiscovery vendors but there has to be a plan for efficiently incorporating them into existing and future product portfolios. Not to pick on Symantec, but they will never be an “eDiscovery company” like Clearwell was. What they can be is a storage management company serving data-intensive business processes like litigation eDiscovery and compliance. Now, that’s useful.

Vendors we are following: [Symantec](#), [HP](#), [IBM](#), [EMC](#), [Oracle](#), [Dell](#)

9. From Backup Tape to Archiving

Archiving is getting more popular for business processes like eDiscovery and I say it's high time. Backups are essential for short term data recovery but were not designed to serve long-term retention needs. Yet many corporations still use backups for long-term data retention instead of deploying archive processes. We believe that archiving adoption is on an upward path but making the capital case for it can be a challenge when backup tape is already a fact of life.

However, there are the strong drivers for adopting archiving over using backup tapes for storage management, eDiscovery, governance and data migration. Backup exists for two reasons – disaster recovery (DR) first and long-term compliant retention second. But companies do not only require DR and provable retention; they also need searchability and recoverability. And without true archives, searching and recovering tapes on a per-file basis presents some serious problems: 1) The long-term cost of transporting and storing tape is not cheap, 2) remediating tape is a big job, and 3) most importantly, *tape offers poor file-level recoverability and searchability*, the two capabilities you absolutely must have for eDiscovery collections. True archiving adds tremendous value to the storage environment by meeting these basic challenges.

Our take: There is a huge cost and time difference between running eDiscovery against archives instead of backup tape. Organizations that already own archiving products need to put more emphasis on additional search and collection capabilities for their archives. Businesses that have not yet invested in archiving can get over the capital purchase hump by understanding the business need of archiving, and by finding archiving solutions that are highly cost-effective and simple to deploy and scale. Fortunately more organizations are recognizing how far superior archiving is to using backup tape as their “archiving” method. Which it is not.

Vendors we are following: [IBM](#), [EMC](#), [Spectra Logic](#), [Metalogix](#)

Analyst [Christine Taylor](#) specializes in eDiscovery, information governance and archiving for [The Taneja Group](#). For more about our eDiscovery survey results and analysis, please visit <http://tanejagroup.com/profiles-reports/category/ediscovery>.