



# EARLY CASE ASSESSMENT: EVOLVING FROM TACTICAL TO PRACTICAL

by Angela Bunting and George J. Socha, Jr., Esq.

## CONTENTS

Executive Summary.....	2
A Traditional Approach to Early Case Assessment.....	3
The Need for Technology.....	4
Confusing Early Case Assessment with Technology Assisted Review.....	4
The Emerging Value of Analytics in Early Case Assessment.....	5
Statistical Analytics.....	5
Date Visualizations.....	5
Textual Analytics.....	6
Relationship Analytics.....	6
Five Practical Tips for Data Analytics in Early Case Assessment.....	7
Find Out What You Have.....	7
Look for Issues in the Data.....	7
Learn What Your Key Players Hold.....	7
Answer the Who, What, and When.....	7
Reduce the Noise.....	7
About the Authors.....	8

## EXECUTIVE SUMMARY

In eDiscovery, we maintain many bad habits from the pre-digital age. However, we have lost sight of one great habit: a preliminary walk-through of boxes containing material potentially relevant for disclosure. This ritual gave the discovery team a chance to get the lay of the land. The team could get a general sense of the boxes’ contents, begin to understand how the materials were organized, and start to prioritize efforts to understand and handle the materials. Boxes containing interesting documents got moved to the front of the line. Boxes of lesser interest were assigned a lower priority, to be returned to as time, need, and resources permitted. The lowest priority boxes were probably never looked at again—time or money ran out or the team decided they had found enough pertinent materials to meet their needs.

Just as we walked through paper documents in the past, today we should do walk-throughs of electronically stored information (ESI). An ESI walk-through needs to be done quite differently, as it poses challenges and offers opportunities we never faced with paper.

The volumes of ESI involved in discovery today make this task impossible without the use of technology. But many legal professionals seem to think early case assessment (ECA) and technology assisted review (TAR)—specifically, technologies like predictive coding—are essentially the same thing. However, these three-letter acronyms have very little in common. Predictive coding is in fact least suitable for ECA but many other technologies can do a much better job of helping you get an early understanding of your data and what to do with it.

In this paper, we will examine four kinds of analytics—statistical, date, textual and relationship—and five practical ways you can apply them in early case assessment. These techniques can give you a much clearer, more accurate, and comprehensive view of the issues at hand and the potential pitfalls you will need to avoid, at the earliest possible stages of the discovery process.

In this paper, we will examine four kinds of analytics—statistical, date, textual and relationship—and five practical ways you can apply them in early case assessment

## A TRADITIONAL APPROACH TO EARLY CASE ASSESSMENT

Back in the good old days when discovery involved wrestling with a mass of paper documents, lawyers took a very practical approach to early case assessment (ECA). An attorney or paralegal new to the process might use the A-to-Z approach the first time, but probably never again. Roughly speaking, they would assemble all the documents in a warehouse, start at the northeastern corner and work their way to the southwestern corner, opening every box, then every folder and then every document. Usually it only took a couple of boxes before the reviewer realized that this would not turn out well.

More experienced folks took the “Sound of Music” approach. Rather than diving right in to the first document, they figured out the lay of the land, starting with a panoramic view and slowly zooming in for a more detailed view. Slowly the team would go through the boxes of documents looking for indexes that might provide clues of what each box contained and if it could be relevant to this matter. They would take notes about potential issues with documents, such as items stored on floppy disks or microfiche, or oversized architectural plans that might need to be treated by a specialist team. There might also be piles of documents in a different language that would require a translator.

As team members grew to understand the layout of the warehouse, they would always make a map of the documents. Then in future visits to the warehouse, they would know which sections of the warehouse had documents most likely to be of interest to the case, and which they could walk past.

Fast forward to today, and yesteryear’s warehouse of boxes is now a computer, a server or server farms full of data. A single hard drive can contain more content than a warehouse of old information. And yet what do we do with all that information? Somewhere in that transition, we have lost the skills of navigating the warehouse. Too often, we revert to the A-to-Z approach deciding we need to set eyes on almost all documents involved in the matter, just in case we miss something. Perhaps this change has come from not trusting that people would file electronic documents in the same way—and with the same diligence—they used to file paper documents.

Whatever the reason, the volumes of electronically stored information (ESI) are increasing. As communication technology has become more mobile and “always on,” the lines between business records and personal ESI have blurred. Many people use corporate email for personal messages, and company devices double as personal devices when we work away from the office. Further complicating matters, people also access work material using personal home-based and portable devices.

These changes make it necessary to bring efficiencies back to the discovery process. It is time to draw on lessons from the past and we must learn how to walk through the digital warehouse.

As communication technology has become more mobile and “always on,” the lines between business records and personal ESI have blurred

## THE NEED FOR TECHNOLOGY

Although it is impractical to put eyes on every document, ESI does not come with a “master index” to advise us if the contents of each folder might be relevant. To make matters more challenging, experience has also taught us that people rarely name their files with records management principles in mind—a key item could be named “Document1” or “Lunch tomorrow.” Quite simply, we can’t possibly conduct efficient ECA without the assistance of technology.

EDRM defines technology assisted review (TAR) as “a process for prioritizing or coding a collection of documents using a computerized system that harnesses human judgments of one or more subject matter expert(s) on a smaller set of documents and then extrapolates those judgments to the remaining document collection...”

While EDRM’s definition sounds very specifically like predictive coding, there are many technologies that can speed up review, including near-duplicate detection, email threading, and clustering. All these tools can find more documents “like these” by combining a variety of algorithms with a subject matter expert’s judgement and can help us zoom in more quickly on sets of material we might want to examine in detail or even one by one.

In an audience poll during our recent webinar “Early Case Assessment is Elementary, Dear Watson,” participants were much more likely to use email threading and near-duplicates than predictive coding as their primary techniques for efficient ECA (see Figure 1).

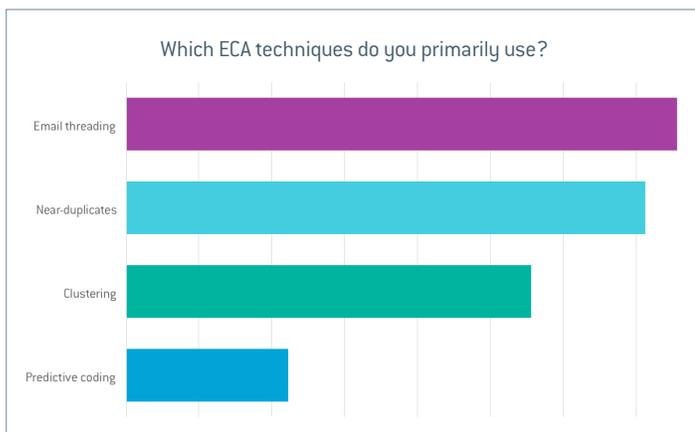


Figure 1: Poll of webinar participants, October 2015.

## CONFUSING EARLY CASE ASSESSMENT WITH TECHNOLOGY ASSISTED REVIEW

At the same time, a majority of discovery practitioners believe ECA and TAR are essentially the same. Most of our webinar attendees (84%) said ECA and TAR were essentially the same thing in our next question (see Figure 2).

It is hard to understand how these two terms came to be synonymous. In fact, there are several downsides to using TAR for early case assessment. Firstly, TAR requires someone to review and understand the documents so they can code email endpoints or create a “seed set” of documents and extrapolate those decisions to other documents in the set. The goal of TAR is to quickly arrive at two sets of documents: relevant and non-relevant. While this approach might save time, it’s hardly the selective browse through the warehouse we discussed earlier. Review can too easily become a process of reinforcing preconceived notions instead of exploring data for information that can support or refute the stories litigators need to build even if technology is assisting.

The other challenge of TAR, and especially predictive coding, is that it relies solely on the textual content of the documents and ignores the metadata. Not only do these systems have little or no ability to deal with video or audio content, pictures, and structured data such as relational databases, it is hard and laborious to manually draw relationships between documents and people simply by reviewing their textual content.

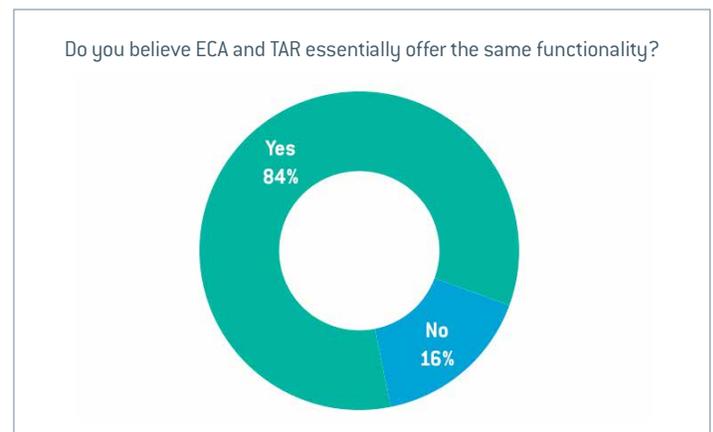


Figure 2: Poll of webinar participants, October 2015.

## THE EMERGING VALUE OF ANALYTICS IN EARLY CASE ASSESSMENT

Analytics is becoming more ingrained in successful businesses as they realize the intelligence they can extract from the data they collect in their day-to-day business operations. However, it has not seen the same exuberant uptake in organizations trying to address their discovery needs. Only a small number of attendees in our webinar said they used analytics that were more sophisticated than a timeline in their discovery process (see Figure 3).

This is unfortunate, because analytics can be the digital equivalent of that first walk-through the warehouse in the days of paper. There are many analytics and visualizations that are worth exploring as part of an ECA workflow. They most often fall into four broad categories: statistical, dates, textual, and relationships.

### Statistical Analytics

Statistical visualizations allow you to look at your ESI from a purely numerical view. They are probably the most commonly seen analytics in discovery. Using metadata and content, you can see a quick summary of your data, identify the top data types in your case, and see if the data set contains languages that require translation or new file types that need the specialist treatment, the way microfilms used to. Pivot visualizations—graphing two categories against each other—are a great way to find anomalies or trends that you might need to address.

### Date Visualizations

Date visualizations very quickly show you trends in your data by making use of the rich metadata and numerous dates that come with ESI. It is a highly effective way to perform a gap analysis of your collected data, looking at outliers and indications of missing data. Combining date visualization with other metadata such as custodian information provides deep insights into what data you have and, more importantly, what you might be missing.

There are many analytics and visualizations that are worth exploring as part of an ECA workflow

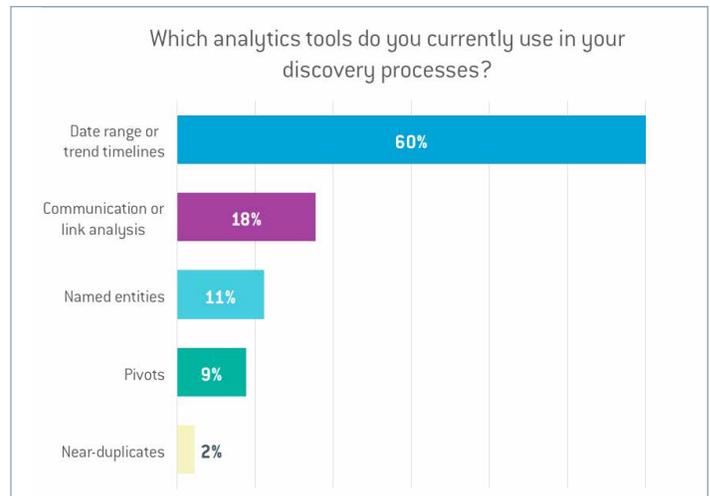


Figure 3: Poll of webinar participants, October 2015.

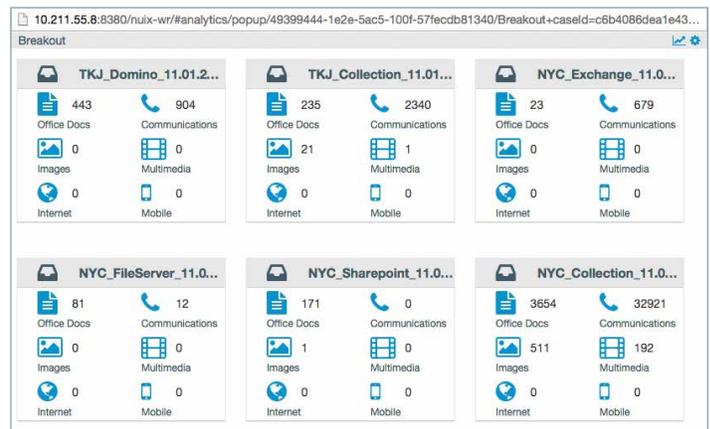


Figure 4: An overview of data sources and their contents.

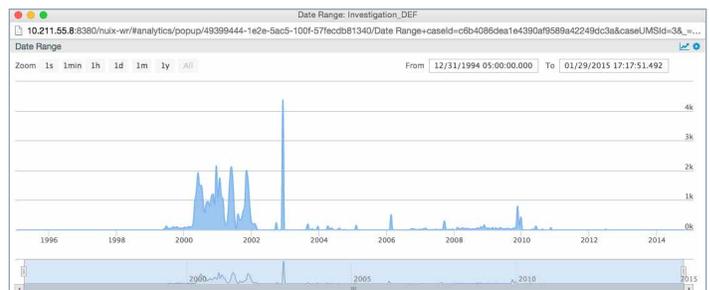


Figure 5: A timeline visualization.

## THE EMERGING VALUE OF ANALYTICS IN EARLY CASE ASSESSMENT cont

### Textual Analytics

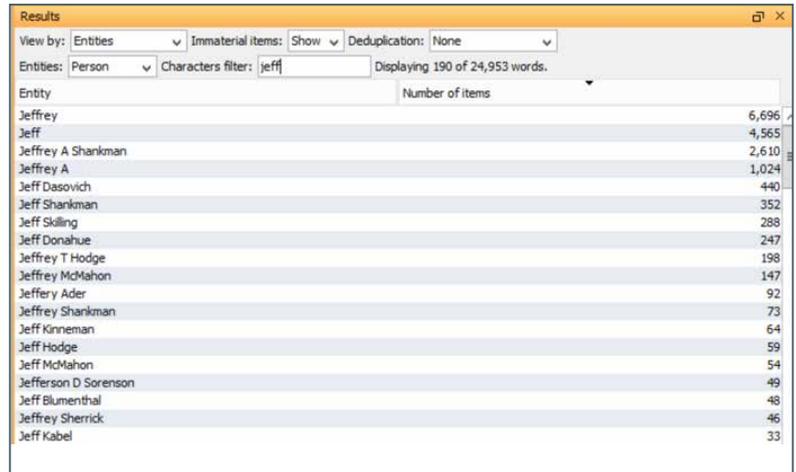
Textual analytics use the content of the documents to provide rapid insight. Discovery practitioners for many years have used word lists—summaries of each word in a set of documents and the number of times it appears. While word lists can be useful to find interesting words and potential misspellings of key terms, they provide little or no context.

The next evolution of textual analytics uses techniques such as shingle lists (sets of sequential words around key terms, such as five words in a row) or topic modelling to provide a better understanding of the context of the words and thus whether they are likely to be of interest to inform the issues at hand. Another interesting new textual analytic is named entities—lists of extracted data based on common patterns of letters and numbers. This makes it easier to identify important items such as social security numbers, credit cards, and client-specific information such as patent numbers or bank accounts.

### Relationship Analytics

Relationship analytics are powerful for analyzing the behavior and associations of custodians. At their simplest, relationship analytics provide a handy map of who communicated with whom and how often, making it easier to pinpoint key custodians or persons of interest including those you may not have originally identified. They can also help to identify inappropriate communications which, if highlighted early, could make the difference between proceeding with a matter or attempting to bring it to a quicker resolution.

Advanced relationship analytics can identify links that are independent of communication information, using common attributes such as files, metadata, or named entities. These can uncover links that might otherwise not be apparent or expected.



Entity	Number of items
Jeffrey	6,696
Jeff	4,565
Jeffrey A Shankman	2,610
Jeffrey A	1,024
Jeff Dasovich	440
Jeff Shankman	352
Jeff Skilling	288
Jeff Donahue	247
Jeffrey T Hodge	198
Jeffrey McMahon	147
Jeffrey Ader	92
Jeffrey Shankman	73
Jeff Kinneman	64
Jeff Hodge	59
Jeff McMahon	54
Jefferson D Sorenson	49
Jeff Blumenthal	48
Jeffrey Sherrick	46
Jeff Kabel	33

Figure 6: A list of “person” named entities and their frequency in the data set.

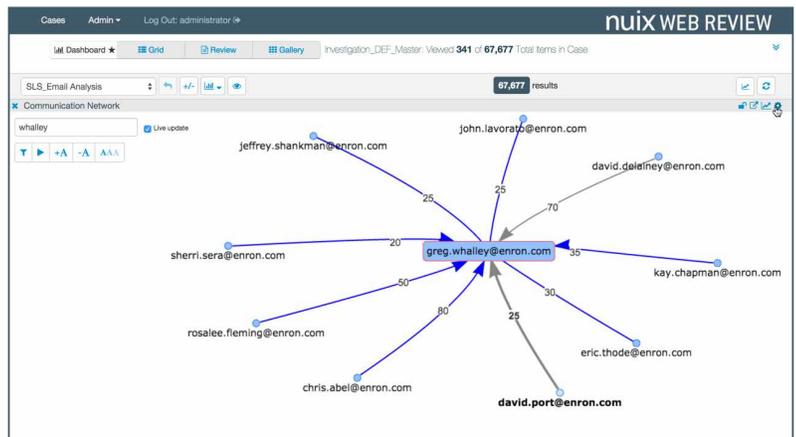


Figure 7: A communication network visualization showing the frequency of communication between custodians.

Advanced relationship analytics can make links independently of communication information, through common attributes such as files, metadata, or named entities

## FIVE PRACTICAL TIPS FOR DATA ANALYTICS IN EARLY CASE ASSESSMENT

Knowing what types of analytics are available gets you halfway to making good use of analytics. Equally important is a practical methodology for applying them. Here are five tips that provide a good starting point to understanding your case earlier.

### 1. Find Out What You Have

Step one: begin exploring your data. Just like that first walk-through the warehouse, an initial stroll through your data gives you an electronic filing map of your ESI. What file types do you have? Can you immediately rule out some folders such as those containing known system files? Will any files require conversion first or need special treatment to transform them into a more easily readable format? What about documents in other languages? Do you think they might hold relevant material that would warrant them being translated?

### 2. Look for Issues in the Data

In a bygone time, items listed on a box's index might not match the contents of the box because they had not been catalogued properly, someone had removed or misfiled documents, or vermin had snacked on them (seriously!). These days, a very common issue that may require further investigation is invalid or anomalous dates. Visualizing data in a timeline and seeing a bunch of items from January 1, 1980 or 2026 is usually a signal that something went wrong at some point between or including the creation and the processing of files containing those dates. Gaps in the timeline might indicate missing data or a shift in IT practices.

Exploring anomalies that might at the surface appear to be issues with the data might reveal other intriguing patterns. For example, an individual who seems to be regularly creating or emailing corrupted files might actually be sending information outside the organization in inappropriate ways or might be changing the file extension to avoid being detected by email scanners.

### 3. Learn What Your Key Players Hold

With paper documents, it was usually easy to find the key documents belonging to a certain project or person. People and organizations kept far less information, and they tended to organize it better. By contrast, ESI tends to be stored all over the organization, often in ways and in places that are not related in any obvious way to the content of the files. Consequently, it may not be obvious if the contents of a particular data source are relevant to the issues you care about.

Textual analytics can provide greater insight into how your key players used the phrases or topics you are interested in. It can reveal the “inside language” of a specific organization, project, or team—and unexpected connections between people, places and things.

### 4. Answer the Who, What, and When

Most disputes hinge on the behavior of the people involved in the issue at hand. Look for relationships and unexpected links which may disprove the involvement of various parties or cause you to question your original theory about the story that unfolded or the scope of the issue.

Going beyond communications, it is often useful to examine links such as shared files or relevant named entities. These may reveal additional relevant people or other information that leads you to change your theory about what happened.

### 5. Reduce the Noise

While analytics is great at quickly understanding the contents of your ESI, it can be just as valuable in helping to cull irrelevant material from your analysis. For example, due to the proliferation of personal communication in corporate systems, it is often helpful to remove spam, personal subscriptions, and even photos and jokes—and analytics can certainly help with that process. Clearing out this digital deadwood will ensure your analyses—such as exploring relationships, topics, and key terms—are more closely aligned to what happened within the issues at hand.

Most disputes pivot on the behavior of the people involved in the issue at hand

## ABOUT THE AUTHORS



### Angela Bunting

Director of eDiscovery Products and Solutions

Angela is one of Australia's eDiscovery pioneers and innovators who has worked with unstructured data and associated technologies for more than 15 years. She has actively participated in advancing the legal technology industry through practice directions and industry associations. Angela's career has followed her passion for driving innovation into the challenges of discovering and producing unstructured data. She is currently responsible for eDiscovery solution development at Nuix and previously worked in managerial and technical roles at Mallesons Stephen Jaques and as the technical lead at Law in Order.



### George J. Socha, Jr., Esq.

Founder, Socha Consulting LLC

George Socha is the co-founder of EDM and the president and founder of Socha Consulting LLC, an electronic discovery consulting firm. George is an advisor and expert witness who focuses on the full range of electronic discovery activities. His clients include corporations, governmental agencies, legal vertical market software and services providers, investment firms and law firms. Before launching his consulting firm in 2003, George spent 16 years as a litigation attorney in private practice.

TO FIND OUT MORE VISIT:

[nuix.com/eDiscovery](http://nuix.com/eDiscovery)

### ABOUT NUIX

Nuix protects, informs, and empowers society in the knowledge age. Leading organizations around the world turn to Nuix when they need fast, accurate answers for investigation, cybersecurity incident response, insider threats, litigation, regulation, privacy, risk management, and other essential challenges.

#### North America

USA: +1 877 470 6849

» Email: [sales@nuix.com](mailto:sales@nuix.com)

#### EMEA

UK: +44 203 786 3160

» Web: [nuix.com](http://nuix.com)

#### APAC

Australia: +61 2 9280 0699

» Twitter: [@nuix](https://twitter.com/nuix)

