edt.

WHY DO LITIGATORS NEED TO THINK LIKE INVESTIGATORS?

The keyword hammer should not be the only tool in a litigator's e-discovery toolbox. Nor is the predictive coding power-saw suited to every case.

One thing that is always needed is investigative analysis, supported by the right technology to uncover a compelling narrative; a story that will win the day for your client.

WHY DO LITIGATORS NEED TO THINK LIKE INVESTIGATORS?

A new client engagement hits the desk. It's a massive litigation project. Of course, time is of the essence. Litigation hold notices issue and a considerable amount of data is rapidly collected.

A quick calculation shows that it would take decades for one lawyer to eyeball every electronic document in the tranche of data that is left after removal of system files and duplicates. Given the amount at stake, that's clearly not going to be a proportionate exercise. Nor is it likely to be effective.

KEYWORDS: THE HAMMER IN A LITIGATOR'S TOOLBOX.

The legal team feverishly starts working on a keyword list to be used to cull the data down - debating the nuances of Boolean "AND", "OR", "NOT" logic, drilling down into the minutia of multiple terminology variations, fine tuning intricate proximity searches, delicately placing wildcards in just the right position within the search strings, hypothesizing the potential application of concept expansion and search term stemming techniques.

When keywords are the only hammer in a litigator's toolbox, all searches look like nails. That means data analysis that is better suited to other approaches often gets distorted into keywords, compromising the potency and accuracy of results. This happens usually because litigators are using the only tool they know how to use or the only one at their disposal. Furthermore, because of the limitations of most software platforms, the whole keyword production exercise is usually performed in the dark before the legal team have access or insight into the data itself.

If that sounds like dangerously wielding hammers in the dark. It is.

In any event, the final compilation of keywords is ultimately issued to the technology team with a simple, blunt instruction to 'generate a hit report'.

Even at first glance, an experienced legal technologist receiving the list knows it's likely to be a futile, painful, exercise.

Boolean "and/or" operators are used inconsistently and incorrectly often delivering the exact opposite result to what was intended. The search terms themselves are so common and patently obvious they will generate hits in practically every document in the dataset ("contract" is probably not a good keyword choice in a contract dispute). Issues are used in place of search terms (in a fraud investigation, the word 'fraud' is probably not going to be within the vernacular of a perpetrator). Person based metadata has been dumbed down into flat, one-dimensional keywords that lose important context information, such as who sent what to whom, and there is at best a futile attempt to accommodate the multitude of possible name variations for key people of interest.

The legal technologist however dutifully follows instructions, obediently applying the search criteria as directed and returning the hit response list.

The legal team discover, to their dismay, that the meticulously crafted search terms will deem more than 90% of the documents to be responsive. Mild panic sets in and some increasingly cavalier adjustments begin. Iterative cycles of anxiously applied search term changes generate successive hit list response reports repeatedly until the *number* of relevant documents is whittled down to what is considered to be a *quantity* of documents that can be reviewed at a cost that is 'proportionate' to the amount at stake.

After this flurry of activity subsides we may, indeed be left with a *quantity* of documents considered to be acceptable. That's fine. But, shouldn't the real question be: Do we have the *right* documents?

IS PROPORTIONALITY FORCING US TO FOCUS ON QUANTITY RATHER THAN QUALITY?

It's all well and good to have a proportionate number of documents but how can we ensure that the documents retained are those most likely to be relevant the dispute or investigation?

Unfortunately, it seems that our increasing focus on proportionality has often forced emphasis on the *quantity* of documents retained for review rather than the *quality* of those documents in terms of their actual importance or relevance to the issues in the case at hand.

There is also a risk that a rote, mechanical application of the proportionality test can, at times, pursuant to the law of unintended consequences, be distorted to justify exorbitantly expensive solutions in high stakes litigation when alternative, more effective lower cost options might actually be available. Perish the thought that high stakes litigation could be resolved fast, at a low cost by finding enough key documents after only a few hours of intelligent legal analysis to support a rapid settlement in a client's favor.

The primary question should not be; how much can I spend on e-discovery as determined by reference to the amount at stake. Rather, how can I deliver the most effective and efficient outcome for my client as the first and foremost consideration and then qualifying this by the obvious requirement to cap expenditure at a level that is proportionate to the issues at stake.

WHAT ABOUT PREDICTIVE CODING?

At a recent industry conference, a well-respected e-discovery thought leader make a somewhat offhand comment to the effect that the e-discovery choice available to litigators hasn't changed for many years in so far as it is still the same old *ho-hum* decision between either keywords or predictive coding.

Really? Is that it?

Is that where our industry finds itself fifteen years on and in this day of powerful, pervasive, data analytics and intelligent technology?

Of course, you could apply predictive coding technologies and methodologies to a project and, as long as the data suits this approach, the budget can accommodate it, the client is willing to take a leap of faith, a senior litigator is prepared to put aside a few days to train a machine, and you have a project manager with experience in this domain, you'll likely achieve some great results.

At the end of the day though, to some extent, when training computers to code predictively aren't we just training them to simulate a bland, one dimensional document by document assessment of relevance? Sure, a computer will do it faster and probably more consistently but where does the intelligent lawyering fit into that scenario? Was it Peter Drucker who said there is nothing more disheartening than doing efficiently that which should not be done at all?

In any event, apart from the keyword hammer and the predictive coding power saw, what are the other options available to a modern-day litigator?

THINKING LIKE INVESTIGATORS.

There are other approaches to discovery. But they require lawyers to think a little more like investigators. That means re-igniting some discovery skills of a bygone era; analysis, creativity, lateral thinking, deep interpretation of the facts and good old fashioned common sense. Skills that were de rigueur before e-discovery flooded our law firms with a quagmire of data and a whole generation of aspiring legal graduates were consigned to toil away on tedious document reviews at the bottom of the e-discovery pyramid.

Armed with powerful and supportive analysis tools, these re-ignited skills will facilitate true investigation of the facts and the issues, exploration of the evidence and development of 'the story' from the very beginning of the case, not just at the tail end when preparing for trial after the pain of e-discovery has subsided.

Here are some considerations.

BUILD YOUR STORY.

At the beginning of the case, it can be extremely empowering to actually put aside data considerations, for a moment anyway, and to think deeply about your client's story.

What is the narrative and what are the compelling themes that will underpin it?

Ask the key questions: Who, What, Where, When and Why.

The answers to these questions will evolve throughout your case as new facts emerge and your knowledge increases, and will provide a cornerstone for your ongoing analysis. They will also, ultimately, provide the platform for the continual refinement and presentation of your story to those who will decide the outcome.

IDENTIFY PEOPLE OF INTEREST (POI).

One of the most important steps is to identify the people of interest (POI) early. Not the custodians. Not yet at least. It's about identifying the witnesses or people who were involved in key events or know the facts surrounding the case and who are likely to be interviewed or deposed. These are some of the questions to consider.

1. Who are the People Of interest (POI) i.e. the people who know, or are most likely to know the relevant facts and circumstances and the other key players.
2. It can be helpful to draw these actors on a whiteboard and to show visually their inter-relationships – alleged, established and contentious.
3. Make note of their roles, titles and other important attributes relevant to the story.
4. Try to articulate clearly what you are trying to show in terms of their communications and interactions between each other.

And **then** it's time to turn to the data. Armed with knowledge gleaned from the exercise outlined above you are well equipped to ask the following questions as you dive into the data.

5. Are there multiple versions of POI names (e.g. Robert = Bob = Rob) that all need to be accommodated when searching?
6. What about initials for middle names, spelling variations or common spelling mistakes?
7. What are their professional email addresses?
8. What are their private email addresses?
9. Could they have used other names e.g. nicknames or code names or false names?
10. Are they connected or interacting via social media networks?
11. Are they interacting via alternative messaging systems?
12. Do we have all the data we need from the POI's including private email and social media accounts?
13. Is it possible that a POI may have communicated via the email of a relative, friend or personal assistant and if so, should the collection net be cast across their data too?
14. How can we normalize the multiple versions of a POI's name appearing in the dataset so that a single easy term can be used for searching throughout the case?

POI'S ARE NOT NECESSARILY CUSTODIANS.

It's important to remember that custodians are not necessarily persons of interest, and vice versa. That is, despite best endeavours, it's quite possible that the data from a key POI may not be available for collection and conversely, it's very common that data was collected from people that are of no real relevance to the key issues in the case. However, such custodians could have sent data to or received data from some of the POI's. So, it is entirely possible to obtain important communications involving POI's even if you have not managed to obtain their own data i.e. even if they are not custodians. This is a commonly misunderstood aspect of e-discovery that often taints the effectiveness of not only collection, but also analysis and review.

GOOGLE IS YOUR FRIEND.

A quick google search surrounding the circumstances of the case, the key people, the alleged facts or events or any other areas of uncertainty can often deliver great insights that are not easily found in the data collected for e-discovery purposes. In cases that have attracted public interest, in particular, it is often surprising to see the freely available information that has been uncovered by journalists who are, after all, professionally trained investigators and skilled story tellers.   So why not leverage their hard work to help build your case? Of course we need to be circumspect about what we read but the point is, litigation is often a voyage of discovery that need not be constrained by the limitations of the discovered documents.

TALK TO YOUR POI'S

In many jurisdictions, there is a whole generation of young lawyers who have not been trained to effectively interview witnesses and who turn instead to mechanically review the documents at the beginning of a case in search of insight into the facts and issues.

Interviewing people is indeed a powerful alternative and a lost art. A quick strategically focused, carefully articulated, early chat with a key witnesses can sometimes provide insights that can save an exorbitant amount of time and expense by illuminating facts that would take many hours to uncover in the digital abyss.

GET TO KNOW THE LINGO.

In interpersonal networks, whether they be a social clique, a sporting community, within a company or an industry, there often emerges over time an idiosyncratic parlance that is reflective of culture, shared experiences, folklore and collective anecdotes. This vernacular can be rich in slang, jargon and double entendre.

It can fast track an investigation to have some insight into this lingua franca before any meaningful searching of electronic data is undertaken.

In addition to this terminology consideration, there is also an inherent human tendency to use code words where there is an intention to disguise intent or true meaning. For example, in one case the word 'chocolates' was commonly used when referring to an incentive payment or bribe – not a word that would normally find itself on a keyword list.

Obtaining insight into these sorts of terms through tools that quickly identify concepts used in unusual contexts or through simply asking a carefully selected POI about the jargon can provide great insights and potentially save a lot of document review time.

USE KEYWORDS ITERATIVELY

If you listen to the hype, keywords have fallen from grace it seems since the advent of sexy analytics and computer assisted review. Fact is, there is still a role for them. They are one of many tools that can be deployed to help a litigator's search for relevant documents. They just need to be used more

effectively. That means, iteratively, real time, with the data at hand so that results of each potential term that is under contemplation can be tested instantly by the legal analyst rather than handballed to a technologist who delivers back a hit report that delivers no insight.

Real time keyword refinement facilitates a quick peek into the actual responsive text within the documents and fine tuning of the terms via repetitive feedback loops until the context hits are delivering the desired results. When developed this way, and used for the right scenario, alongside other analytical tools, keywords can become powerful indeed.

BUILD YOUR FACTS AND CHRONOLOGY OF EVENTS

If you start with an initial understanding of what is alleged to have happened, when these events occurred and the people who were involved, this provides an ideal framework for exploration of data throughout the discovery process.

Traditionally the chronology and story building exercise takes place after the discovery process and during a trial preparation phase. However, this case management paradigm if embraced at the outset, even before the collection commences, will provide a canvass upon which you can weave the rich tapestry of your story as the case progresses. You can position the characters, the facts and the sequence of events such that the story unfolds perfectly by the time it reaches maturity and is presented for resolution.

A COMPELLING NARRATIVE WILL WIN THE DAY.

Once you have the framework for your story, you can certainly use keywords, where that makes sense and of course, you could contemplate predictive coding too. But please treat these as merely two of many tools and analytical techniques at your disposal.

All modern e-discovery or litigation support software platforms have functionality to support the approaches outlined here. This is not about rocket science or artificial intelligence. We are talking about *real* intelligence - common sense, inquisitive and creative legal analysis.

What really matters is the story. Your creativity, your insight, your intuition and your incisive legal analysis will help you to develop and tell the story. At the end of the day, it's a compelling and credible narrative that will win the day for your client so it makes sense to start developing it early using the right tools and the right data.