

## TechTalk:

---



### Accelerating Review in an Imperfect World

By Kevin Carr, President, InterLegis

*Editor's note: Each month, Kevin shares his insights about discovery technologies in an easy-to-read straightforward manner. Check each edition of ALSP Update to stay abreast of his view on technology topics and trends. In this article, he discusses the effective ways legal teams can accelerate document reviews when dealing with large volume, complex discovery collections.*

---

In a perfect world, we would load up raw data into a computer, input the background of the case and it would spit out the documents we need to win the day. But we do not live in such a world. In fact, given the dynamic, complex nature of the data we deal with in e-Discovery, I would say we live in a grossly *imperfect* world.

The perfect world scenario mentioned above is the proverbial "Holy Grail" of our existence – discovering critical information as quickly (and cost-effectively) as possible.

To do that, we necessarily have to move from the time-consuming and expensive process of linear review strategies and migrate to a streamlined, cost-effective and *accurate* process of dynamic or non-linear review. During the last four years, we've made significant progress toward that goal; however, most review teams are not yet fully streamlined in their processes. The industry is still evaluating a combination of new technologies and developing best practices.

As data collections continue to grow and become more complex, we need to make substantial progress toward reaching this "perfect world" we so desire. Because the reality is that the total volume of data collected for the average case makes it physically impossible to perform a total linear review – we have no choice but to embrace advanced technologies to assist us in our efforts.

As Seen in:



So how do we get to the next level of accelerating review in a meaningful, accurate and defensible fashion? This is the question I continue to ask myself each and every day. To answer that question, I've categorized solutions into four areas.

Accelerating review involves:

- Technology
- Processes
- Reporting
- People

## Technology

As with any growing industry, as new technologies are introduced to the market, we begin to refine them into a combination of useful processes. So much has happened in the e-Discovery technology space compared to just a couple of years ago. We are developing cutting-edge solutions that bring artificial intelligence and advanced analytics right to our fingertips. This allows us to quickly uncover critical documents in a fraction of the time it would normally take to perform a linear review.

Below is an overview of the advanced technologies we can employ to accelerate review. Many you know ... some you may have heard of ... and some may be new to you. But even for those that you're familiar with, we are finding new advancements in how to use older technologies combined with newer ones. I could describe each of these in great detail, which would stretch this article into a mini-novel ... and it's already long enough! I'll just hit the highlights:

### 1. Data Mapping:

As the name implies, these technologies are the best way to quickly comprehend where (and how) data lives to understand what you have. When dealing with a vast array of data types, the right data mapping technologies can quickly plot out of number of things at each stage of the discovery life cycle including:

- **Collection** – Where does data live on a network, archive or specific computer? Have we collected the right data?

.....  
*"We are developing cutting-edge solutions that bring artificial intelligence and advanced analytics right to our fingertips in ways that allow us to quickly uncover critical documents in a fraction of the time it would normally take to perform a linear review."*  
 .....

*(Data Mapping – Continued)*

- **Early Case Assessment** – Do we have the enough data to support the case? Should we settle or move forward?
- **Processing/Culling** – Quickly identify relevant data to “normalize,” index, and push to reviewers.
- **Review** – Quickly identify and make document decisions.

## 2. Culling Technologies: Full Attribute/Metadata Filtering

The right culling technologies will significantly speed review and drastically reduce costs. Users can effectively weed out irrelevant data, which means there is less data to process and fewer documents that need to be reviewed. Therefore, with less information to deal with, fewer dollars are spent. Good culling technologies should do more than just allow users to search through the data. The data should be available to you *before* you fully process it. The tool should map out all the “moving parts” of data collections, including text and all metadata. It should offer advanced analytics such as concept groupings, similarity matching, relationship mapping, and more. And finally, it should allow you to make categorical decisions on which data is ultimately relevant/responsive.

.....  
*“The right culling technologies will significantly speed review and drastically reduce costs.”*  
 .....

## 3. Visual Analytics: “Leveraging the Visual Nature of the Human Mind”

This is one of the most effective technologies you can employ to accelerate review. The most effective of these technologies can show data in an illustrative format so that all the important (or not so important) information contained within can come to life. As opposed to the traditional method of working down a linear flat list of search results in order to read and understand documents individually, visual analytics can help you quickly identify spikes in activity, activity trends, communication threads, relationship between important selected attributes in the data set, and make categorical decisions on the collection. What used to take many months can arguably be fleshed out in a matter of minutes.

.....  
*“What used to take many months can arguably be fleshed out in a matter of minutes.”*  
 .....

*Examples of visual analytic tools:*

- **Clustering** – Allows you to visually group data together that share common attributes. And, to maximize effectiveness, this technology should do more than just group by concepts. You should be able to cluster *ALL* attributes in the database (custodians, file types, authors, dates, and more). Example: "Show me all of Bob Smith's documents that are e-mails and contain discussions around the concept of 'contract negotiations.'" Powerful stuff!
- **Timeline/Spikes of Activity** – Because nearly all litigation is date-specific, it goes without saying that having the ability to plot out all data activity on a timeline is invaluable. This allows users to quickly hone-in on spikes in activity and recognize relevant activity trends. Example: "Why was there a dramatic increase in documents dealing with the concept of 'contract negotiations' from 2 May to 14 May?"
- **Correspondence Mapping** – This gives reviewers the ability to quickly map out who is speaking to whom. This is especially useful for e-mail conversations, but certainly does not have to be limited to that. We often find that the basis of the litigation can be found in such communication between parties, so why not simply map out those communication threads to quickly uncover those events? With the right technologies, you should be able to limit those communication threads by other data attributes, such as: concepts, date ranges, selected organizations, domains, and more. Example: "Show me all the communications discussing contract negotiations on the Widget Project between certain individuals at ABC Corporation and XYZ Limited within the first two quarters of 2007." With this sort of specificity at your fingertips, you will quickly find highly relevant information to the case.
- **Relationship Mining** – Often, it's the "common threads" between data that flesh out relevant documents. Therefore, having the ability to select which attributes you are interested in and then map out everything that shares those relationships is a great way to quickly uncover important documents. Example: "Show me all e-mail or Word documents associated with Bob Smith, that contain the concept 'contract negotiations,' and were created within the first two quarters of 2007." And just like that, the precise documents you need rise to the top!

.....  
*"The right visual analytic technologies can show data in an illustrative format so that all the important (or not so important) information contained within can come to life."*  
 .....

*(Examples of visual analytic tools - Continued)*

- **DataBurst** – If you wanted to quickly see all the attributes that exist within a collection, a DataBurst graph would be beneficial. With DataBurst graphs, users could select either the entire collection or filter it by specific attributes – such as by custodian or other combinations – and then quickly pick documents that contain the selected attributes that are interesting. Example: “Of Bob Smith’s documents dated from January 2005 through December 2007, show me all concepts, file types, and recipients that are contained within.”

#### 4. E-mail Threading and Analysis:

With the bulk of business communication happening via e-mail, the ability to maintain complete e-mail threads is essential to accelerating review. And, with the right technologies, there’s a lot of incredible analytics that can examine and evaluate on e-mail collections.

*Examples of e-mail analytics:*

- **Complete E-mail Thread** – The basic functionality of being able to see the complete thread to make a single decision about the entire set.
- **Last E-mail Identification** – The ability to validate the last e-mail in a thread. The benefit of doing so means that reviewers only need to look at that single e-mail and read backwards to see the full story. Of course, this only works if the technology can validate that the complete thread is present in the last e-mail (see next bullet).
- **Change in Threads** – There are many benefits to having the ability to identify if a thread is broken or otherwise truncated.
  - **Last e-mail validation:** In the case where you are looking for the last e-mail in the thread to see the complete conversation, it would be helpful to know if any of the story is missing via intentional deletion by the author when composing the last e-mail, or if the sender or custodian’s e-mail program has truncated some of it due to settings or size.

.....  
*“With the bulk of business communication happening via e-mail, the ability to maintain complete e-mail threads is essential to accelerating review.”*  
 .....

*(Examples of e-mail analytics - Continued)*

.....  
*"We're just scratching the surface on what's possible with e-mail analytics ... "*  
 .....

- **Change in recipients:** The ability to see if any recipients were left off or added to the thread at some point. Once this sort of change can be identified, reviewers can try to figure out why this happened and if it's relevant to issues of the case.
- **New thread creation:** Have you ever seen an e-mail and realized you want to send a new e-mail to every recipient? So, what do you do? You simply take the original e-mail, hit "reply to all," delete the body and subject, and start a new e-mail. However, all of the underlying information about that new e-mail is still related to the original thread. Identifying when this sort of event happens could lead to relevant communication or prove that certain parties did in fact receive that original e-mail.

- **Similar/Related Threads** – The ability to select what attributes are relevant in one e-mail thread (names, domains, dates, concepts, attachments, etc) in order to find others like it.

- **Timeline Analysis** – And of course, the ability to quickly plot out e-mail threads on a timeline to see when all this activity took place.

.....  
*"Good similarity matching technologies should be able to "read and understand" all documents in a database to quickly group those that have similar attributes."*  
 .....

We're just scratching the surface on what's possible with e-mail analytics, and continue to make many time-saving advancements to this type of technology.

**4. Similarity Matching/Profiling:**

This describes the ability to take a single document (or group of documents) and ask the technology to "find others like it/them." This goes well beyond duplicate or near-duplicate detection. Good similarity matching technologies should be able to "read and understand" all documents in a database to quickly group those that have similar attributes.

*Examples of similarity matching/profiling technologies:*

- **Straight Similarity Matching:** This technology finds similar documents for you by evaluating various attributes inherent to the document such as: author, recipient, date, subject matter/concepts contained within, page formatting and more. In a single step, you will have the ability to pull all similar documents together and make corresponding decisions on them all. This type of technology should provide you a list of documents ranging from duplicates, to near-duplicates and ultimately documents that have a large percentage of the attributes listed above in common.
- **Selected Attributes Similarity Analysis:** A slight variation on the above, in which you can indicate which specific attributes are important to you so that the technology can place a higher weight on your selections and return results based on those weightings.
- **Document Profiling:** This technology allows users to create a fictitious document that contains all the attributes any relevant document should contain. For example: a user could create an e-mail by a selected author, select a list of recipients, with a specific date range, containing certain text or concepts, with certain possible attachment types. Once this "template" is set, the technology goes out and finds all that share those similar traits and show the results back in order of relevancy. Furthermore, when additional documents are added to the database, the technology can proactively notify case administrators when new matches are found.

## 5. Concept Analysis:

We're all pretty familiar with concept technologies by now. However, like most technology out there (in our industry or not), we're only scratching the surface as to what is possible.

*Here are some recent advancements in concept analysis technologies:*

.....  
*"A major issue with concept technologies is that they have traditionally employed a "one-size-fits-all" approach to their analysis. "*  
 .....

- **Matter-Specific:** A major issue with concept technologies is that they have traditionally employed a "one-size-fits-all" approach to their analysis. However, when dealing with matters that deal with specific industries or terms of art, it is important to have concept analysis technologies geared towards that specific data set. What means one thing to the pharmaceutical industry could mean something completely different to the manufacturing industry.
- **Foreign Language:** We've made great progress with analyzing English-based concepts, but advancements in the ability to decipher concepts within foreign languages are becoming more and more critical.
- **Customized Concept Analysis:** This describes the ability to "program" the technology as needed to ensure it helps you find what you need. For example, you can associate a company's product name with: a specific company name, certain individuals within an organization, or certain project code names (i.e. the "Phoenix Project"), departments/locations (i.e. Toledo facility), certain terminology/industry lingo, with distribution channels/clients and much more – the sky is the limit. The benefit is the ability to custom fit the technology for the matter at hand. Once done, you can release the technology to "read and understand" the data they way you would and return highly relevant intelligence as to what is contained in the database.

.....  
*"Making more than one document decision at a time is essential to accelerating review."*  
 .....

**5. Categorical Marking/Dynamic Review:**

Making more than one document decision at a time is essential to accelerating review. This functionality is a major component that moves review from a linear environment to a dynamic one. However, effectiveness is maximized by using the right technologies capable of clustering documents together by similarity, common threads, or concepts. To speed review even further, try categorical marking using visual analytics in order to map out trends in data and make on-the-fly decisions on the graph itself based on what you see.

## 6. Automated Workflow:

This is the next big area of technological advancement, and describes the ability to set up certain rules or criteria that will automatically move a document (or groups of documents) through to the next stage in the discovery life cycle. Typically, this "next stage" is reviewer assignment.

### *Examples of workflow technologies:*

- **Auto-Batch Assignment:** Simply taking new batches that hit the review system and equally distribute them across reviewers or assign documents based on certain data attributes.
- **Rules-Based Auto-Assignment:** Based on various criteria or events that take place (usually how a document is coded), documents will automatically be assigned to the next reviewer in queue.
- **First-Pass, Second-Pass:** Similar to above, it's typically used to validate privilege or responsiveness.
- **Profiling:** Instead of auto-assignment via issue code rules, the right technology can allow you to set up various document profiles by selecting a combination of attributes that indicate specific relevancy. By doing so, the technology seeks out all documents that fit this profile and assigns them to the right reviewers who specialize in this type of data.
- **Auto-Production:** The use of issue code auto-assignment or profiling to automatically mark documents for production. Obviously, this technology should be used with sampling QC and validation technologies to ensure accuracy (see below).
- **Foreign Language:** The ability for technology to recognize foreign languages and assign those documents to reviewers that can deal with that specific language.

.....  
*"Automated workflow is the next big area of technological advancement."*  
 .....

*(Examples of workflow technologies - Continued)*

- **Area of Expertise:** The ability for technology to identify certain subject matter, usually based on case-specific concept analysis, in order to assign documents to reviewers who are focused on a particular area of expertise. Example: pharmaceutical, aviation, biochemistry, patent law, etc.

.....  
*"... native redaction technologies allow legal teams to see where a redaction is needed, make the redaction right then and there, and move on ... resulting in huge time savings."*  
 .....

## 7. Native Review / Native Redactions:

We're all aware of a major benefit of native review: no TIFFing, which saves time and reduces costs. However, new advancements are expanding the types of documents that can be viewed within native viewers.

One of the biggest advancements we have developed is the ability to redact native documents on-the-fly ... and do so without having to first TIFF the document. And even though it might take a minute or so to convert a native file to TIFF (assuming background TIFFing is available in your review application), those minutes add up significantly in large document reviews. Therefore, native redaction technologies allow legal teams to see where a redaction is needed, make the redaction right then and there, and move on ... resulting in huge time savings.

## 8. Sampling/Quality Control:

Sampling can be used to accelerate steps along the discovery life cycle:

- **Collection:** This ensures the right collection processes are being employed. This helps legal teams avoid having to collect all of *what they think* is the right data and pay to process/analyze all of it – only to find out they've haven't collected correctly. Instead, legal teams can use sampling to pull in a small amount of data, analyze it, validate it's the stuff they're looking for so they can go out and collect with confidence.

*(Examples of Sampling/Quality Control – Continued)*

- **Culling:** Validates culling criteria (keyword or otherwise). Are we getting a large number of false positives, resulting in irrelevant data, and why? Instead of uncovering the answer to this question by unnecessarily processing and reviewing *ALL* irrelevant data, sampling at this stage can greatly reduce e-Discovery processing expenditures.
- **Review:** Either validates dynamic document coding criteria or, as a last step, protects privilege or verifies responsiveness.
- **Production Quality Control/Validation:** Ensures a number of things before final release of productions, including: protecting privilege, confirming redacted text, error-free TIFF conversion/branding, proper load file creation and more.

### 9. Combination of the Above:

Of course, to maximize productivity, you'll need the ability to easily tap into any combination of these technologies. No two cases are exactly the same. For example, some, based on volume or scope, may lend themselves better to conceptual analysis. Or some are better suited for visual analytics. The point is, you need to have an arsenal of tools at the ready so that you can react to what you see in the data and make on-the-fly decisions on how to best accelerate your review. And to that end, if you can have all these tools readily available at your fingertips in a single application, the more productive your review will be.

When we talk about accelerating review or making the discovery process more efficient, it is easy to focus on technology. After all, it's cool. It's intriguing. It can boggle the mind as to how the technologies work. And, it's fun to play with all the new whiz-bang tools that get introduced to the market. Once you find the right combination of technologies (based on your project needs), they truly can speed the culling, processing, review and production of your discovery data.

.....  
*"Once you find the right combination of technologies (based on your project needs), they truly can speed the culling, processing, review and production of your discovery data."*  
 .....

*However*, there are additional pieces of the puzzle that are just as important as technology. **Without proven processes, reporting and people, the technology is useless.** This is no different than buying an expensive sports car ... only to find it is missing the steering wheel!

.....  
 "Without proven processes,  
 reporting and people, the  
 technology is useless."  
 .....

Maybe the following topics aren't as sexy as things like: artificial intelligence, concept analysis, e-mail threading, similarity matching, and visual analytics ... but hopefully you will find new insights in how to strategize for your next case.

## Processes

As mentioned, having the right technologies is only part of the equation. To put these tools to work for you in the most effective manner, you need sound review *processes* in place. Here are some things to consider when developing your review strategy:

### 1. Take a Funnel Approach:

In order to ensure clearly irrelevant or non-responsive documents are immediately weeded out at the top of the funnel so that the smallest, most relevant data can be output from the bottom of the funnel into review or production. The idea here is that, during the first step of the process, it is critical to "trim the fat" from the collection by analyzing all the moving parts of the data in order to carve out that which is truly relevant. Because data collections continue to grow in massive proportions, it is more important than ever to move from identifying what is *likely* relevant to that which is *highly* relevant. To do this, sound culling technologies (as mentioned above) designed around effective funneling processes are critical.

### 2. "One-Look" Philosophy:

This describes the ability to see a document (or better, a group of similar documents) as early as possible in the discovery life cycle – ideally at the culling/processing stage – and make a single and final decision on what to do with it (issue code, produce, cull, etc.) so that it doesn't have to be reviewed again. Of course, some documents need validation as part of a second-stage process. However, the more documents you can review with only "one-look" ... even if it is as little as 10–20 percent of the database ... the more you will realize tremendous time and cost savings further down the review process.

### 3. Dynamic Workflow:

In order to truly streamline review, legal teams must use technologies coupled with processes that allow for dynamic review — as opposed to linear review. Average document volumes are simply too large these days to have reviewers evaluate each and every document in a database. Instead, legal teams should first group or cluster documents together by concepts, common threads, or unique document relationships in order to make decisions in chunks. Typically, review happens in multiple stages so that first stage reviewers are using dynamic technologies to make categorical decisions so that second stage reviewers can validate those decisions and take action from there. For example, the following can be decided during a dynamic first stage review:

- What is potentially relevant?
- What is clearly *irrelevant*?
- What is possibly privileged?
- Which documents should be routed to subject matter specialists?
- Which documents contain foreign languages?
- ... and more, based on the issues of the case

### 4. Linear Review has its Place!

There's a lot of importance placed on dynamic or categorical review, resulting in the notion that linear review is a bad thing. As a technologist, I'm a big believer that you should put advanced technology to work in a dynamic environment not to *eliminate* linear review, but to *reduce* the need for linear review and reduce costs in the process. The goal here is to minimize the volume of documents that review teams must evaluate one-by-one. However, the next step in the dynamic workflow described above should be to validate those grouped decisions through **linear review**. Ideally, the goal should be to deal with 80–90 percent of data in dynamic fashion in order to flesh out the 10–20 percent that should be reviewed linearly.

### 5. Trust in (Tested) Technology:

Of course, none of this works without solid technologies to drive the process. However, there must exist absolute trust in the technologies you are using. Therefore, it is critical for legal teams to fully test these tools before putting them to work. Find out how these technologies work, how long they've been in the market, what other success have they had in the past, what their pros and cons are, and what are some of the underlying methodologies that make them do what they do. Lastly, be sure to use technologies that can be fully supported 24/7 by the provider of these tools. Proper and ongoing training and strategic consulting should be provided to ensure you are maximizing their effectiveness.

## Reporting

How do you know just how effective your technology and processes are? How can you demonstrate the defensibility of your processes? Reporting is the key to answering these questions and more.

Comprehensive reporting can facilitate the following areas in various ways ... just to name a few:

1. **Collection:** What's been collected? From which custodians and locations? How has the data traveled through the discovery life cycle and who has been responsible for it (chain of custody)?
2. **Culling/Processing:** What is the inventory of all data? What are the exceptions and errors in the collection (uncommon file formats, proprietary formats, password-protected files, corrupt data, etc)? What has been culled out and by which criteria? What is the total volume import vs. export?
3. **Review:** Review assignments. Total documents per day/hour. Total review status. Reviewer productivity reports. Reviewer decision tracking report. Unreviewed documents. Privilege logs. Workflow assignment criteria results. Ad hoc reporting based on case needs and production reports.

## People

And of course, none of this is possible without the right people in place to manage these processes and use these technologies. I've written in the past about how critical people are to this process (ALSP TechTalk, August 2008).

The best technologies are completely useless without:

- Competent people who understand the discovery process and the dynamics of the case.
- Comprehensive training so that users have a solid understanding on how to most effectively use these tools.
- Responsive and 24/7 support from technology providers who stand ready to answer questions about using these tools.
- Project Management by both review teams and service/technology providers.
- Strategic Consulting in order to assist legal teams in devising the best use of technologies and workflow, as well as the ability to validate the defensibility of these tools.

## The Quest for "A Perfect World"

So, as you can see by the sheer length of this article, there are many moving parts to developing the perfect solution to truly accelerate review and dramatically reduce costs. And the reality is, what works for one case may not work the same for another. Therefore, a solid understanding of all available technologies – coupled with the right consultative support – is critical in creating the right solution for each matter at hand.

One key to success is having all these tools readily available in one application or by a single provider. This way, as new issues arise during your review, you can make on-the-fly strategic decisions on how to maximize your success. After all, it is asking a lot of legal teams to always make the absolute best decisions on review strategy without first getting into the data and understanding what issues you need to address.

.....  
*"... of course, none of this is possible without the right people in place to manage these processes and use these technologies.."*  
 .....

.....  
*"... there are many moving parts to developing the perfect solution to truly accelerate review and dramatically reduce costs. "*  
 .....

Clearly, the quest for this perfect world is an ongoing challenge. We'll never see the day where you can load up documents into the "hopper" of artificial intelligence technologies, hit the "enter" button, give the machine time to analyze the data ... and out comes the specific documents you need to carry the day. However, as the e-Discovery industry continues to mature, we are finding very powerful tools and developing best practices that can get us as close as possible to this utopian solution.

'Til next time ... KC



.....  
*Kevin Carr, president of InterLegis Inc., has a wide range of Internet-based technology expertise. As the architect of the InterLegis system, he has developed cutting-edge discovery technologies and best practices relating to data mapping, conceptual analysis, electronic data culling/processing, similarity matching, streamlined document review, automatic categorizations, visual analysis, native review, document digitization, optical character recognition, compression, database indexing, advanced searching and document security.*  
.....