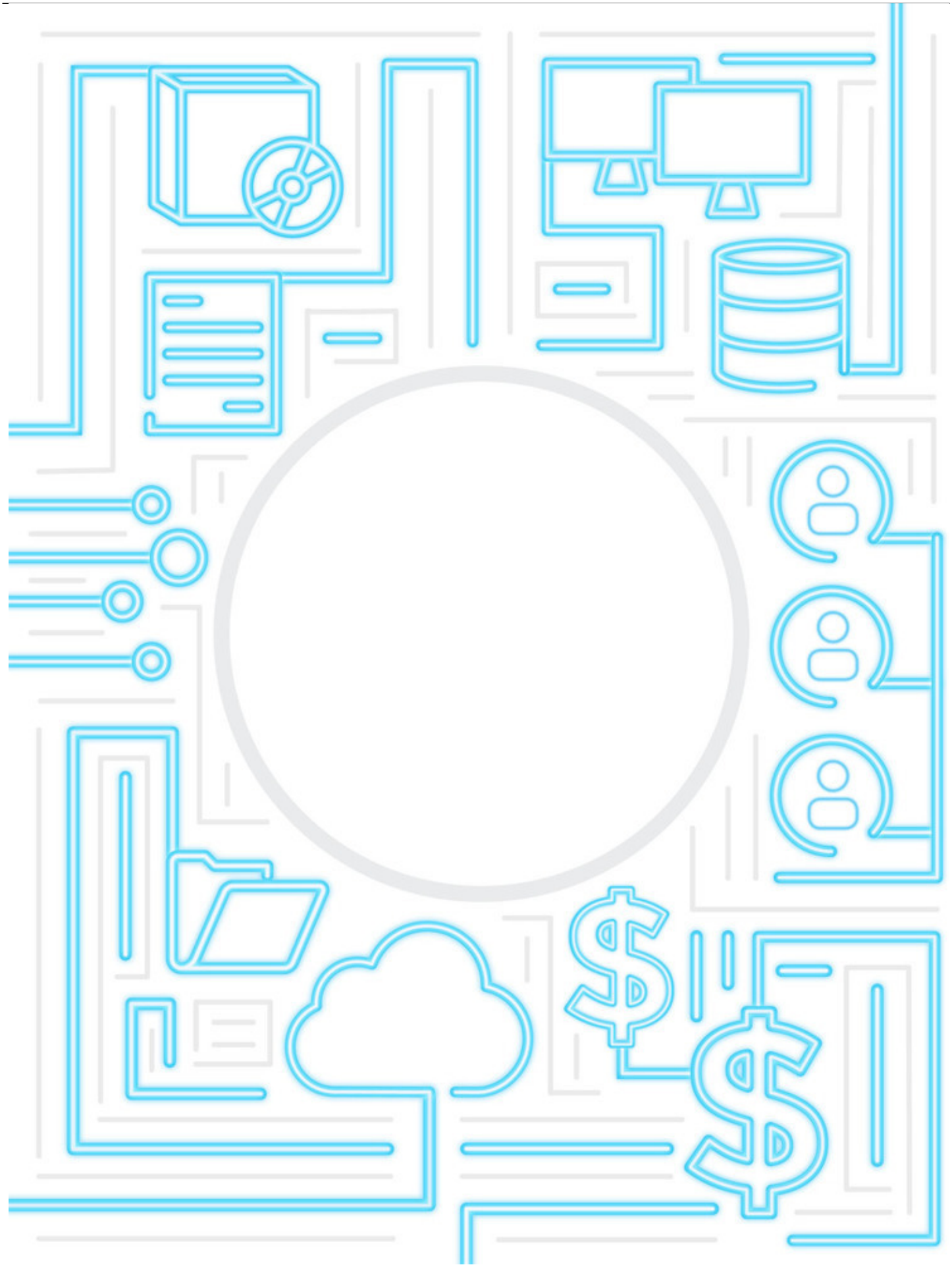
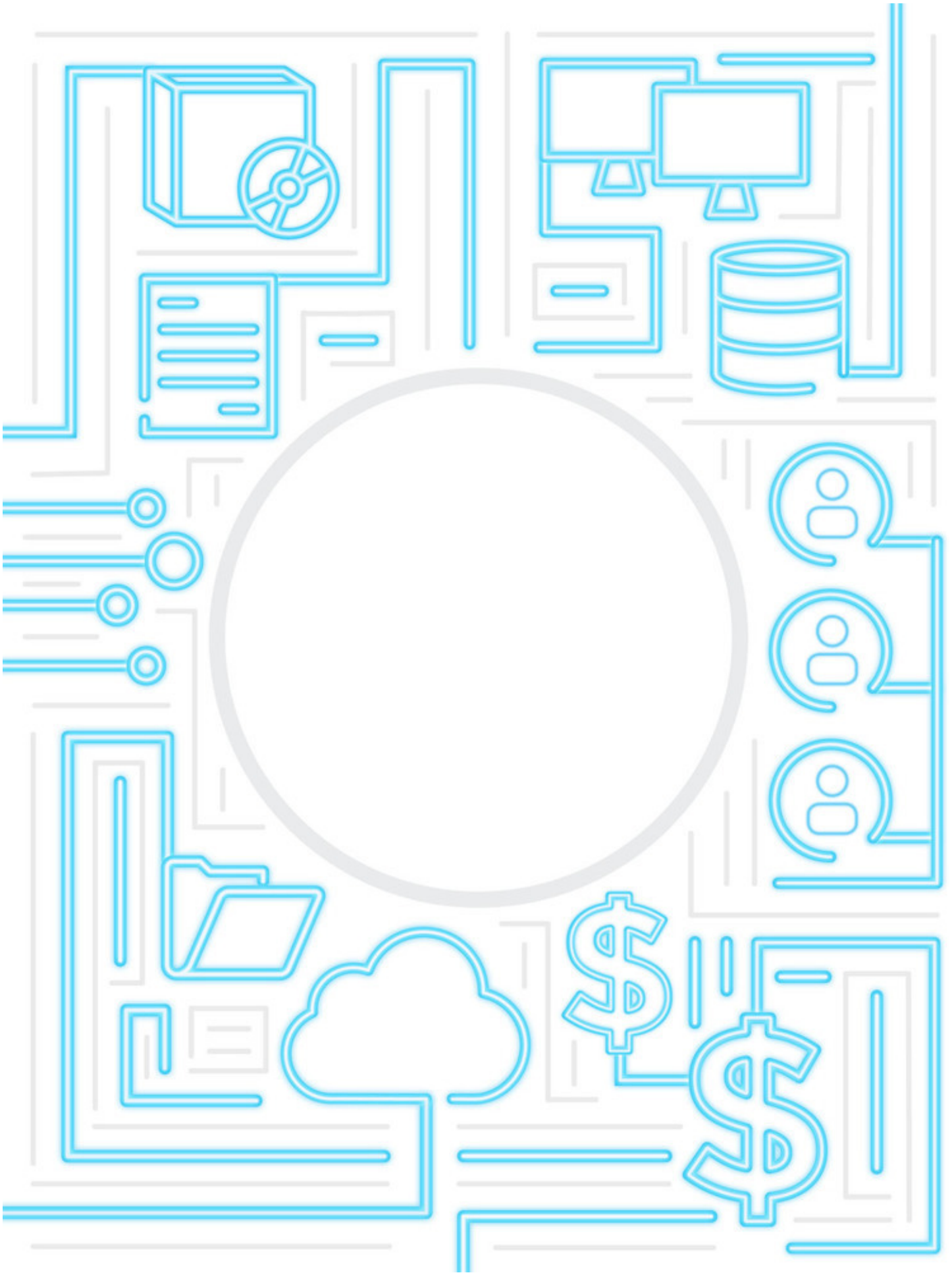
ACC DOCKET
INFORMED. INDISPENSABLE. IN-HOUSE.

Developments in Litigation Technology: Early Sleuthing and Scoping

Litigation and Dispute Resolution

Technology, Privacy, and eCommerce





CHEAT SHEET

- ***Sleuth to the heart's content.*** Due to decreases in pricing models and increases in available technology, early case sleuthing and scoping have become more affordable and feasible.
- ***Decide with more.*** Lawyers can use emails and electronic documents during case scoping to understand what information is available to them before deciding whether to settle or proceed to negotiations.
- ***Share the process.*** If opposing counsel has any questions, explain the strategy and process for locating documents through a screen share meeting.
- ***Avoid tunnel vision.*** An overreliance on technology assisted review to solve discovery related issues can lead to two types of tunnel vision: tool focus and over-emphasis on review.

In litigation, the side that gets the earliest and best understanding of the underlying facts and how the evidence supports or undercuts key witnesses has a distinct strategic advantage. Fortunately, several trends are converging that make early case sleuthing and scoping more affordable and feasible than ever for cases of all sizes.

CLOUD STORAGE

is increasingly inexpensive and highly elastic. Large volumes can be processed on demand for literally pennies per gigabyte without infrastructure investment. Geo-specific cloud storage can keep data in country, avoiding many General Data Protection Regulation (GDPR) problems.

SECURE BANDWIDTH

has increased. Large sets of data can be moved quickly, securely, and economically without physically shipping computer media.

ANALYTICS SOFTWARE

is becoming more ubiquitous. Corporations don't have to load email and documents onto expensive final review platforms to have comparable functionality, e.g., predictive coding, concept wheels, email threading, deduping, text search, etc.

PROJECT-BASED VIRTUAL TEAMS

can come together just for the life of a project and include the expertise needed for any part of the project (e.g., database discovery or dealing with legacy applications). There is no need for large, ongoing headcounts.

SCREEN SHARING

applications permit quick, spontaneous meetings and information exchanges. Litigation teams can be assembled globally and can be much nimbler than only face-to-face meetings.

PRICING MODELS HAVE SHIFTED FROM THE OLD PER-GIGABYTE OR PER-DOCUMENT MODELS

Experts now provide technology at either no additional cost or may bill for direct out-of-pocket cost for things like Amazon Web Services hosting or processing fees — minor expenses compared to final review platforms.

> **As any trial lawyer will say, the story of how the process worked is much more defensible than any measure of recall or precision through sampling, which will always be less than 100 percent.**

These trends offer opportunities and solutions with distinct advantages over traditional methods of document review and predictive analytics.

Early sleuthing and scoping

Being able to collect, process, and analyze email and electronic documents for little or no cost allows lawyers to understand what information they have when deciding whether to settle or to negotiate on a preliminary scope for e-discovery. It also provides metrics to use in such negotiations.

The truth is that if corporations use people who understand robust software, they can find whatever they need in just about any collection of documents, without a lot of cost and heartache. With the early investigation and scoping approach to e-discovery, outside counsel will be reading only relevant documents for intelligence and understanding, not just to make “in” or “out” relevancy decisions or apply “confidential” or “privileged” tags to documents.

Because early scoping is affordable, attorneys can do this as soon as the demand letter arrives. They can know before the case is filed what happened, who said what to whom, and whether to fight or settle. For every pleading and subpoena, in-house counsel can quickly and inexpensively scope the task at hand by collecting the email for two or three highly relevant custodi¬ans and quickly see with whom they discussed the topics at issue. This information can be used to not only negotiate the scope of discovery with the government or adversary, but also to inform you about who else really should be included as a custodian, based on the electronic evidence, not just supposition.

After the work is done and the relevant documents are produced, any questions by opposing counsel as to completeness can be quickly resolved with a screen share meeting with the lawyer who did the work, explaining the strategy and process for locating the documents. As any trial lawyer will say, the story of how the process worked is much more defensible than any measure of recall or precision through sampling, which will always be less than 100 percent.

Examples

Here are some scenarios of how this new approach to technology-assisted awareness and processing can work.

Products liability — eye and skin irritation

You have a potential product liability situation where employees of a customer have complained

about itchy eyes and skin from using the product at workbenches. Investigators from the National Institute for Occupational Safety and Health (NIOSH) have requested documents, and your client is worried about lawsuits. Under traditional “wait and see” litigation discovery practices, discovery would have been put off because it was expensive.

Basic e-discovery cost control

Here are some basic things that can be done to greatly lower costs without compromising quality. More information on the first three items is available in the ACC Docket article, “Ethics and E-Discovery Review,” Jan./Feb. 2010, pp. 46-57:

DeNIST

- As soon as possible, identify and exclude files that are created by and obtained from software providers as part of their software installations and updates. There is no evidentiary value in these files. The National Institute of Standards and Technology maintains a Software Reference Library that lists these files and their hash values that can be used to identify them.

DeDUPE

- Don’t have multiple instances of the same files repeatedly reviewed by different people. It’s wasteful, and you risk inconsistent production decisions.

THREAD EMAILS

- Emails and attachments that are part of the same email conversation or thread must be tethered together, so that they are read together by one person. This provides the overall context of the whole conversation and avoids inconsistent coding or designations.

BE TRANSPARENT EARLY

- If sweeping discovery doesn’t make sense, get on the phone with opposing counsel, and tell them what you’re planning to do. Use screen share software to show examples of what you’re talking about.

GET A DIVORCE FROM PER GIGABYTE PRICING

- Just as there is growing recognition that companies are better served by flat billing as opposed to per gigabyte pricing, e-discovery should be purchased on a largely flat-fee basis. As Jeffrey Carr has said, buy the right thing.

USE BENCHMARKS

- Participate in ACC and other forums for exchanging information about what recurring tasks should cost. Use that information in negotiating price.

Safety and Health (NIOSH) have requested documents, and your client is worried about lawsuits. Under traditional “wait and see” litigation discovery practices, discovery would have been put off because it was expensive.

With today’s technology offerings and project-based teaming, you can collect email and documents from your key employees without incurring costs and process them into a document repository with investigative analytics and sophisticated search. This could cost as little as US$20 per gigabyte, per month, with no additional fees for generating TIFFs, deduping, threading emails, or persistent highlighting of keywords and productions.

Lawyers skilled in using the platform’s software would investigate the documents collected and provide you and your client with a “who’s who” or “key players” list, a chronology, and meaningful intelligence regarding what your client knew, or should have known, about irritated eyes and skin from using its products in a matter of days. Total professional fees at US$200 per hour are less than US$5,000.

E-discovery proficiency quiz

Lawyers who can’t make full hands-on use of the panoply of tools available to analyze electronic discovery are like dentists who can’t personally use drills. They lose the ability to interact with the data and quickly gain insights from it. Here are some basic tasks that lead investigators or attorneys should be adept at handling.

LEAD ATTORNEY/INVESTIGATOR

Each of the following tasks other than the keyword testing should take less than 10 minutes.

TASK	SIGNIFICANCE
DOMAIN-SPECIFIC USER NAMES. List all user names associated with a specified domain name and the number of emails to/from each user.	This task is critical to obtaining a basic understanding of who the key players are and what the relationships are among them.
“PRIVATE” EMAIL ADDRESSES. Identify all usernames associated with “private” (i.e., non-corporate) email accounts.	Private emails are sometimes used to circumvent monitoring of corporate accounts. E-discovery practitioners should be able to identify the leading “private” email providers without being told (e.g., Gmail.com, Hotmail.com, Yahoo.com, Comcast.com, and Verizon.net).
LIST NAME LABELS. Identify all the name labels associated with a given email address (e.g., the name labels associated with jsmith@enron.com might be “John Smith,” “J. Smith,” “Jack,” or “JS.”)	Name labels are added by people who create contacts in contact lists. There can be as many name labels as there are people who have a given email address as a contact. Listing name labels helps identify nicknames or alternate forms of names, which sometimes appear without the email address or domain for subsequent analysis.
RECIPIENT DOMAINS. Identify all domain names that a specified person has sent emails to or all emails between certain individuals and/or organizations.	Without the ability to see organizations or people with whom certain individuals of interest have been communicating, it’s hard to see overall communication patterns.
COMMUNICATION FREQUENCY OVER TIME.	In litigation involving key dates, a spike or gap in

Present a frequency analysis of the number of emails a specified person has sent or received within a given time period.	communications patterns can be indicative of underlying behavior.
PRIVATE CLOUD STORAGE. Find all documents mentioning online data storage domains used for storing or exchanging data	Cloud storage providers like DropBox, Google Drive, Cloud Drive, OneDrive, OneBox, etc. can be used to store data off corporate networks. Identifying where those terms appear in a data collection can be a way of identifying potentially additional data sources. The person doing the searching should be able to identify these domains/product names without being prompted or informed.
FALSE POSITIVE KEYWORD TESTING. Conduct keyword analysis to identify search terms that retrieve false positives.	Every data set is different and “dirty” search terms that unintentionally retrieve false positives can inflate costs enormously.

DOJ price fixing investigation

You are dealing with a collection of 2.5 million documents queued for review in response to a Department of Justice (DOJ) subpoena regarding alleged price-fixing, and you are well aware that there will be subsequent civil litigation. You have service providers proposing techniques, including various forms of Technology Assisted Review (TAR), for prioritizing the documents so that those of lesser value are reviewed by less expensive lawyers in India. There are also proposals to test the results with sampling and measures taken of recall and precision.

It sounds great but it also sounds expensive, to the tune of several hundred thousand dollars per month, for many, many months. However, there are lower-cost alternatives now possible using cloud computing and lawyers skilled in using sophisticated investigative software. Assuming 10,000 documents per gigabyte, the collection you are dealing with is likely at least 250 gigs. At US$20 per gig, per month, the cost to process, host, and maintain the documents would be US$5,000 per month. In addition, using the techniques mentioned in the accompanying “E-discovery Proficiency Quiz,” such as domain name, file type analysis, and email name grouping, together with quickly finding and removing completely irrelevant and useless files (system files), the volume could be quickly reduced by as much as half — 1,250,000 documents and 125 gigs.

Organizations processing e-discovery data

TASK	SIGNIFICANCE	BENCHMARK
Ingest the electronic files located at an indicated URL.	Nimbleness is a key part of using technology to gain strategic advantage through earlier and better awareness. Lawyers can’t get ahead of things if they can’t quickly start analyzing the	1 hour

	underlying data.	
View and assess exceptions in the ingested file (i.e., things that couldn't be indexed).	Lawyers should be able to look at and assess documents that did not process because of password-protected/ encrypted files, corrupt/unreadable files, and image-only files. Usually, these files are not relevant, and it would be a waste of money to try fixing the issues and processing them.	1 hour
Identify and remove files on the NIST list of standard software files (provided by NIST in their Software Reference Library).	Lawyers should be smart enough not to process program files distributed by software providers like Microsoft and Adobe. Otherwise, charges are inflated and the associated data collections are cluttered with items with zero evidentiary value.	10 minutes
Describe what happens to data when the case is over.	It is very important that the lawyers and/or consultants you use have a case closing checklist, similar to the one provided in an earlier *ACC Docket* article "The Case is Closed. Where Are Your Documents?" May 2013, pp. 57-67.	10 minutes

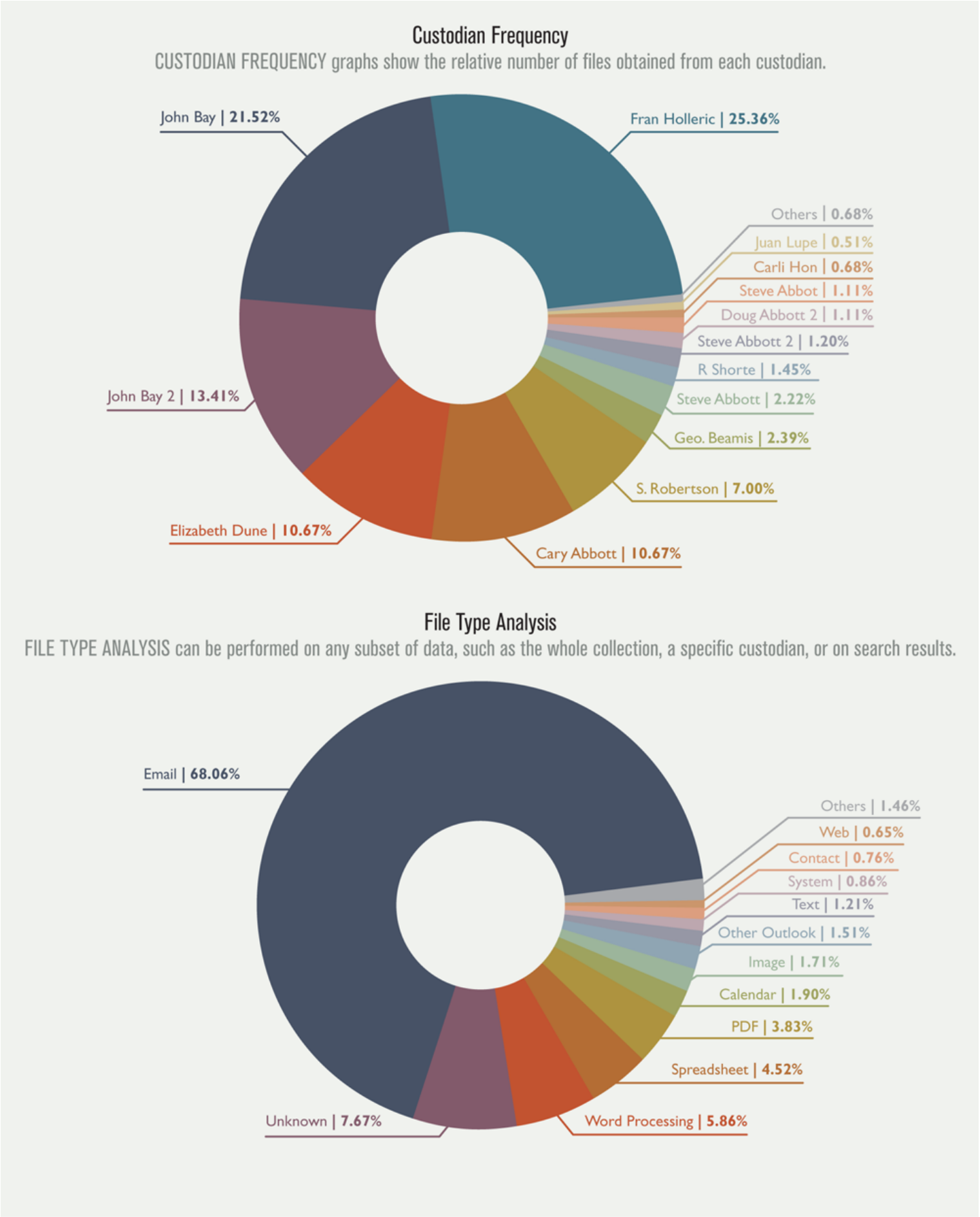
In an elastic cloud environment, this means that your server needs are reduced by half and your monthly costs go down to US$2,500 per month. With today's document analytics tools, it wouldn't take much to find out from the documents who was involved in making pricing decisions, their emails about pricing and who they were sent to and received from, and who they may have met with from competing companies.

Two lawyers skilled in using the software's analytics capabilities, with an understanding of the allegations and key players both within and outside the organization, could likely identify the relevant documents in 200-300 hours. At US$200 an hour, that is US$40,000 to US$60,000. After the production is made and the DOJ voices skepticism, those same skilled lawyers who found the documents will screen-share with the Assistant US Attorneys and show them how and why they know they found and produced the responsive documents.

Interactive sleuthing and scoping tools

These are five of the types of interactive displays that can be used for early sleuthing and scoping

prior to putting content in a final review platform:



Ten steps for early awareness and relevance/non-relevance selection

There is no single way to gain awareness and select responsive documents. Here are steps often recommended by Jeff Johnson, a Kansas City e-discovery consultant, depending on the circumstances of each case. These steps are typically applied prior to loading content on the final review platform.

1. **SPAM AND NON-RELEVANT STANDARD EMAIL REPORTS AND MESSAGES.** Review emails sent to large numbers of recipients and remove non-relevant messages.
2. **CONCEPT CLUSTERING.** Cluster documents based on textual content — it may be possible to include or exclude large clusters of documents based on examining just a few of them.
3. **NO-RESPONSE EMAILS.** Significant emails tend to occur in email threads with replies and forwards. Emails without responses can be clustered so non-relevant emails can be removed.
4. **INCLUSIONARY SEARCH TERM TESTING.** Based on what has been learned so far, test initial key terms and logic for responsive documents across key custodians, noting related terms.
5. **EXCLUSIONARY SEARCH TERMS FOR NON-RESPONSIVE ITEMS.** Perform iterative non-relevant key terms analysis, and remove non-relevant documents.
6. **CAL**. Use continuous active learning TAR as way to pinpoint key documents prior to sending files to review platform provider.
7. **SPECIFY EMBEDDED OBJECT TREATMENT.** Evaluate whether to create separate “documents” for each embedded object (e.g., a spreadsheet graph embedded in a Word document). “Exploding” embedded objects to create additional documents can clutter the review database and inflate data ingestion and storage costs. The review collection should be audited to ensure the specifications were followed.
8. **LARGEST SIZE FILES.** Examine the largest sized files to avoid paying excessive data loading and storage on final review platforms (e.g., identify large PDFs, graphics, and video).
9. **SENTIMENT ANALYSIS.** Review documents that contain highly charged emotional content (e.g., profanity). Labor or employment cases and contract disputes often involve angry or accusatory emails that help prioritize and scope review tremendously.
10. **VISUAL CLASSIFICATION AND GLYPH SEARCH.** Visual classification technology can cluster visually-similar documents without using or requiring textual analysis. It is ideal for some collections and serves as a cross-check on text-based tools for any collection. The same technology enables searching for key graphical elements, such as logos, stamps, and graphics.

Additional scenarios

Here are additional scenarios and how they might pan out using a strategic approach that leverages lawyers skilled in using state-of-art document analytics software in an elastic cloud computing environment.

- **Claim evaluation.** You receive notice of a claim. By identifying principal players in the matter and doing a preliminary collection and processing of their email, you can quickly get a preview of how the case will play out. The costs to collect and ingest the email for a few of the individuals involved and to find out what happened for many cases, is less than US$1,500. As

noted above, some consultants using Amazon Web Services do not charge for collecting or processing data, instead charging hourly for technical skills in using the software and finding the meaningful emails quickly (in some cases within a few hours).

- **Key employee departures.** A key employee suddenly leaves. Collecting and processing their email can be done for no cost and, once it is in a cloud-based repository and used by someone skilled in its search features, the departed employee's email can be quickly organized and managed as needed, for less than US$2,000 in most cases.
- **Subpoena responses.** Responses to information and document subpoenas that might have run up costs and fees over US$20,000 can now be responded to for less than US$5,000, including processing, hosting, review, tagging, and production.

Scoping the extent and validity of litigation claims provides huge strategic advantages in knowledge and reduced cost to the great benefit of the corporation. Best of all, it is affordable and practical.

Potential TAR tunnel vision

An over-emphasis on TAR (a.k.a. predictive coding) as a way to solve discovery related issues can lead to two types of tunnel vision.

- **Tool focus.** Focusing on the predictive coding type of TAR (classifying sets of documents based on classifying a subset of them) can lead to ignoring other proven tools like concept clustering, domain name analysis, social network analysis, or advanced search techniques.
- **Over-emphasis on review.** Focusing on the review phase can cause lawyers to overlook opportunities for analysis and understanding much earlier in the litigation process. Document review may not take place for many months, if not years, after the suit was initiated, and that is simply too late to go hands-on with the documents. To the extent that corporations find TAR to be a useful way to gain understanding, they should consider using it before documents are placed in a final review platform. They can also use it to evaluate not just outgoing productions but productions from other parties as well.

Conclusion

Scoping the extent and validity of litigation claims provides huge strategic advantages in knowledge and reduced cost to the great benefit of the corporation. Best of all, it is affordable and practical.

Seth Eichenholtz

In-House Lawyer and Head of Electronic Discovery

Mastercard

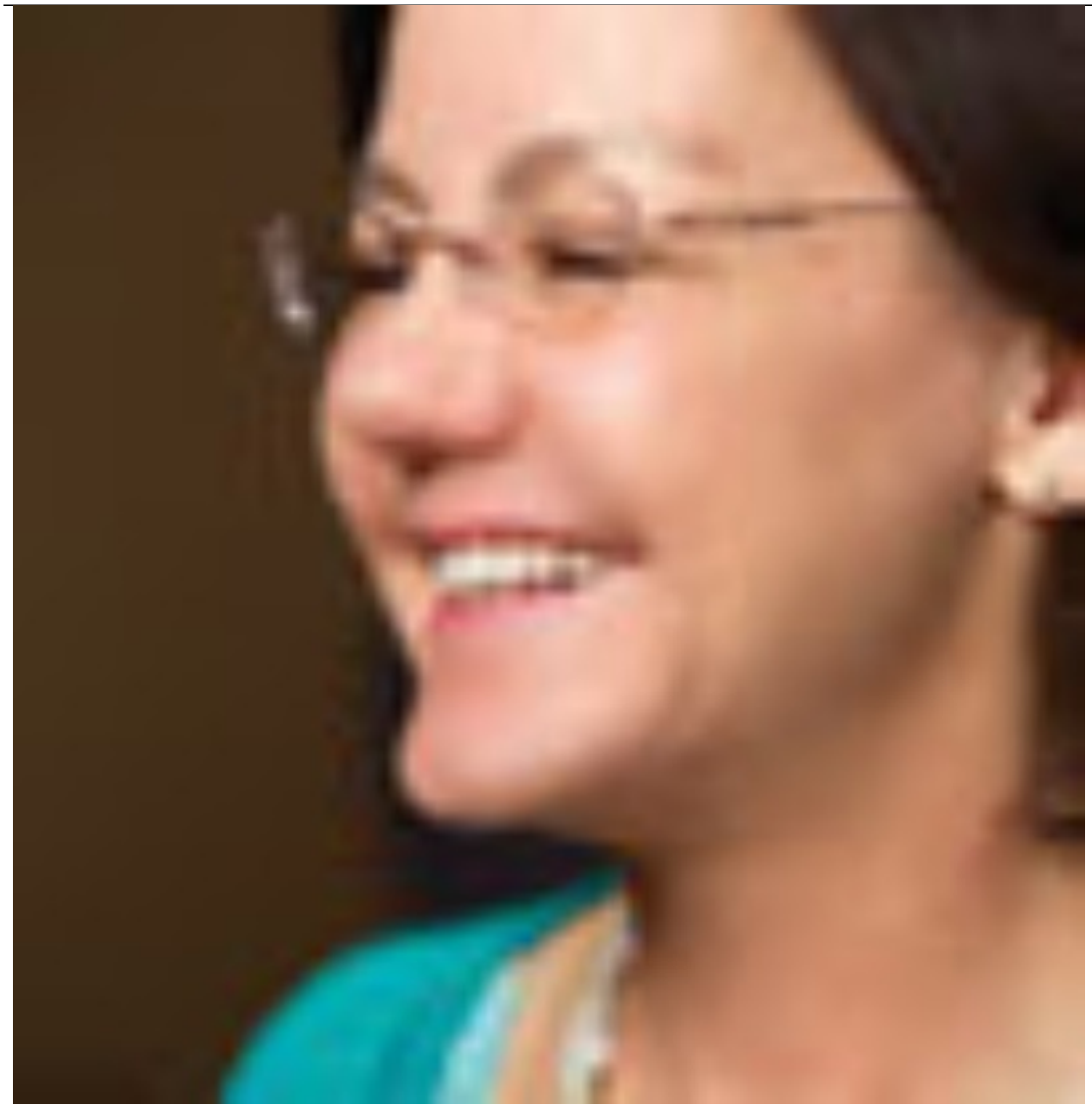
Prior to Mastercard, he worked as an e-discovery consultant and managed e-discovery at Swiss Re.

Tim Donovan

Intellectual Property Attorney

Donovan is an A-rated intellectual property attorney who has held a number of positions as general counsel in large and small corporations in the technology and healthcare sectors. He is currently a legal advisor to a healthcare software vendor in Kansas City and an advisory board member for Advance Law.

Anne Kershaw



Lawyer and Consultant

Anne Kershaw is a lawyer and consultant who has been immersed in eDiscovery for many years, serving as proportionality counsel in numerous cases. She co-authored the Judges' Guide to Cost-Effective E-Discovery, teaches at Columbia University, and has written earlier articles for the ACC Docket on discovery topics.