



Common Problems With E-Discovery—and Their Solutions

ANDRES HERNANDEZ

Discovery is all about being able to track down information that is pertinent to your case. A few decades ago, this meant getting your hands on paper documents, but of course, that's not true anymore.

As our lives and business dealings became increasingly digital, so did discovery. It moved from filing cabinets to computer servers, but unfortunately, the change was not simple or easy.

Even as e-discovery became more and more crucial to building legal cases, it increased in complexity, and a number of common issues have started to make themselves known. Here are the most common e-discovery problems that you are likely to face as well as solutions for how to make the process run more smoothly.

There's Just Too Much Data

You have probably read reports about how much digital data is out there and how fast it is growing. Most, though, have a tough time putting the numbers into perspective. What, really, does 4.4 zettabytes look like versus 44 zettabytes?

Cyclone Interactive tried to put it into perspective in a 2014 infographic,¹ using the memory in tablets from the year 2013 (which, of course, has only increased in the time since). They explain that if you put all of the digital information that existed in 2013 (4.4 zettabytes) into tablets and stacked them up, they would reach two-thirds of the way to the moon. By 2020, that data is estimated to grow to 44 zettabytes, which would result in 6.6 such stacks.

Obviously, you will not have to go through even close to that amount of information on a single case. Doing so would be impossible. But the tiniest slice of a zettabyte is still a stunningly high amount of data.

And sifting through massive amounts of data is no longer a problem solely for mass tort and class action cases. More and more often, relatively small cases—say, for example, a breach of contract case worth a few hundred thousand dollars—are eliciting e-discovery requests that might end up yielding tens of thousands of non-duplicative items. Getting through so much data takes an enormous amount of time and manpower, not to mention cost.

All of us—companies and regular people—have a lot more digital data these days, and it's only going to get worse. When attorneys have to delve into that data to litigate a case, everything takes longer, becomes more complicated, and costs more.

Solution: If you know exactly what document you are looking for, but the main problem is the sheer amount of data provided, you actually have an advantage. Traditional Boolean searches, like you use on Google or Lexis-Nexis, can turn up what you need. This is particularly true if you know the exact file name, but it's also the case if you know specific keywords in the target document. The more you know, the better.

Data is Everywhere

The most common source of e-discovery data is email, but it is far from the only one out there. Today, e-discovery might include data from a variety of digital sources, such as: email systems, social media accounts, collaboration systems, real-time communication systems,

work computers (desktops and laptops), home computers (desktops and laptops), work-issued mobile devices (including tablets and smartphones), employee-owned mobile devices (including tablets and smartphones), corporate blogs and wikis, file servers, USB storage devices, and other repositories of both structured and unstructured data.

There are issues involved at every step along the way. First, you have to track down the data from all of these disparate sources. Then you have to gain access to it. And once you reach that point, you still have to sift through it—potentially mountains and mountains of data—to determine what, if anything, is worthwhile. But you also do not want to ignore a potential data source because it could end up netting you a piece of information crucial to your case.

Solution: There are many ways to collect data, so you should consider the pros and cons of each strategy to determine what's best for your particular case.

Data Collection

“Self-collection” means allowing the data custodians themselves to copy the files on a storage device. This is generally the least reliable option, since individuals may not be technically savvy, could make mistakes, or might overlook documents. Courts have also questioned whether this collection method constitutes a defensible e-discovery response. But it can be cost-effective if it is a small matter, involves

“Remote collection” means integrating data sources with a centralized internal collection system, allowing data to be collected remotely. IT professionals are a part of the process, but don't directly interact with the data sources. This means the process can be completed quickly and efficiently. This strategy involves an upfront investment and deployment process, however, so it is best suited to organizations with regular collection demands.

Not All Data is Created Equal

Beyond the fact that digital data is continually growing and lives pretty much everywhere, there's another issue: some types of data are inherently more difficult to look through.

What we're talking about is structured versus unstructured data. The simplest possible way of explaining these two types of data is that structured data is easy to search using simple operations, while unstructured data isn't. In other words, some data sources are easy to search through for specific terms and information, while others require you to pretty much look at each individual digital document to ensure you're not missing anything important.

Want some examples? Spreadsheets make for good “structured” data because the information is in a relational database system that computer programs can quickly and easily scan. But emails, social media posts, texts, and other phone data are unstructured.

“IT collection” means that members of the IT department collect the data. Unlike individual data custodians or general employees, they have a better understanding of the data landscape, as well as the technical skills necessary to ensure the data is collected properly.

a low volume of data, involves only highly conventional data, and the opposing party and judge have agreed to the plan in advance.

“IT collection” means that members of the IT department collect the data. Unlike individual data custodians or general employees, they have a better understanding of the data landscape, as well as the technical skills necessary to ensure the data is collected properly.

But at smaller organizations with limited internal IT resources, the data collection process can be burdensome for maintaining regular business functions. Also, IT professionals tend to provide forensic imaging versus logical copies when gathering data. This is important to note because forensic copies include additional information, such as previously deleted data.

This can result in too much data being collected, driving up e-discovery costs. And unless there is a suspicion of data tampering, a logical copy will usually meet the court's expectation.

“Third-party collection” means allowing an external professional to collect the data. This is beneficial because it has a smaller impact on the day-to-day functions of the organization. And the third party will have expertise in the procedures, tools, and skills to collect data in a way that will stand up to judicial scrutiny. The downside is that it comes with a larger price tag.

For example, if email was truly “structured,” it would be arranged by content and exact subject, in addition to being organized by date, time, and who sent it. But that's not how people communicate. Even if an email conversation is relatively focused, it is rare that we only speak about one subject per email or that we don't use synonyms or slang terms. All of these things complicate a search and make it more difficult to unequivocally say that certain files are useful—or not—without looking at them one by one.

Solution: There are a number of different ways that you can sift through unstructured data depending on what you know coming into the search such as:

- **If You Know Exactly What You're Looking For...**

Use metadata analysis. You can group documents together based on topics that you know you want to look for and “drill down” into those classifications to get more and more specific. Using email as an example, you could categorize things by: To, From, CC, originating domain, when the email was sent, and so on.

This type of search can be particularly useful if you are searching for information where a topic is being discussed using a variety of different words that may include euphemisms, code names, and other

forms of nontraditional language. In other words, the way people often talk or write.

Use textual analytics. Another way to search is to feed a particular document that includes the type of information you're looking for into the system and have the program search for conceptually similar documents and data. One of the best things about this type of search is that, if you don't have a "control" document with the right kind of information, it's possible to simply create your own—even using multiple documents with "good" data to do it—and tell the system to look for documents like that.

• If You Are Trying to Fill in Knowledge Gaps...

Use keywords intelligently. A cornerstone of e-discovery review is coming up with search terms to define the documents that you believe you need. Many lawyers don't go any farther than that—but they should.

Keywords can be used not to just provide you with a "data dump," as it were, but also to prioritize what you should review. One way to help you in this prioritization is to look at the keyword hit count report. This will tell you how many times specific keywords show up in the documents that you have requested, which can help you determine which files are likely to be the most relevant to your search. Keep in mind that just because a keyword appears a lot does not mean it will actually be useful, but this approach can help you with that issue as well.

How so? If you see a keyword turning up a lot in documents that end up not being helpful, you can eliminate that keyword and try different searches to fine-tune your approach.

And finally, it can be helpful to turn on keyword highlighting in the documents, so you can more readily scan and focus on those areas where keywords appear. This can really speed up the review process without completely eliminating files, and it's especially helpful when looking at documents in groups or families.

• If You Are Still Trying to Understand Your Case...

Use concept-clustering. This analytics tool searches for words and concepts that appear together frequently and creates a report telling you what they are. For example, let's say you are litigating a class-action personal injury case. If you receive a set of documents about the injuries, it is possible that some of them will be talking about the nature of injuries, while others might be arguments over whether a particular injury occurred, and still others will discuss who is at fault.

What a concept-clustering tool does is separate the documents by the concept or content rather than group them all together because they share a keyword. It's a great way to keep documents organized and have a better idea of what you're looking at, as well as providing you with a way to eliminate groups of files if the concept doesn't apply to your case.

Use word frequency hit count. A keyword hit count report tells you how many times the keywords you've chosen appear in individual documents. A word frequency hit count tells you how many times all words appear.

Why is this useful? Because when you begin the e-discovery process, chances are good that you won't necessarily know all the terms that are going to be important. For example, what if you are litigating a case about a particular project and that project is often referred to by a code word or slang term? Running a word frequency hit count

can help you to find those documents where a project may appear under a different name than the one you knew originally.

Use TAR. Technology-assisted review, or TAR, can cut the amount of time you spend on review by up to 75 percent.² To put that in real dollars, let's say your e-discovery costs for a case are \$100,000. According to estimates, \$70,000 of that would be due to review. Three-quarters of \$70,000 is \$52,500. So using TAR could potentially save you over \$50,000—or more than half of the total e-discovery price tag in this instance.

Seems worth it, no? So what exactly is TAR? Think of it as a combination of textual analytics and concept-clustering. Here's how TAR works in a nutshell: You select documents that are valuable and relevant to your case. TAR looks at the content of these documents and performs a search to find similar documents. The more documents you select as relevant, the better TAR becomes at finding the types of data you need.

Some attorneys worry that TAR will eliminate the need for lawyers altogether or prove indefensible in court because they are not manually looking at documents. These fears, however, are largely unfounded. TAR will not eliminate lawyers, but free them up to advocate for their clients.

Moreover, assisted review still requires attorneys to establish the process and evaluate the outcome. You'll still be involved, but you'll no longer need to wade through file after file.

And as for TAR being indefensible, the old "eyes on every document" process has already been determined to be unduly burdensome. Translation: using TAR should be just fine.

It's Just Too Expensive

This has been touched on numerous times under other points herein, but it's worth discussing in-depth. Because large-scale e-discovery used to be the purview of mass tort and class-action cases—in other words, cases with big budgets and even bigger payouts—the pricing model for handling e-discovery needs was largely developed for these types of situations. In other words, it's expensive. Ridiculously expensive. Prohibitively expensive.

Let's say that the collected data in your case equals about one computer hard drive. Not even a big one—500 GB. Of that, there's about 100 GB of reviewable material. To house that data over a year, a typical e-discovery company will charge over \$160,000. Factor in that housing data tends to be about 30 percent to 40 percent of the total cost and 100 GB of e-discovery data will cost you about a half million dollars.³ For small and medium-sized firms handling small and medium-sized cases, that's just not possible.

Solution: Luckily, low-cost options have started to appear, and the overall pricing structure is changing and adapting to meet the needs of both larger and smaller firms. Some examples of cost-effective solutions include:

- Acrobat Legal Edition: www.adobe.com
- CasePoint: www.legaldiscoveryllc.com
- Cicayda: www.cicayda.com
- CloudNine: www.cloudninediscovery.com
- CS Disco: www.csdisco.com
- Digital WarRoom Pro: www.digitalwarroom.com
- dtSearch Desktop: www.dtsearch.com
- Harvester and SafeCopy: www.pinpointlabs.com
- Intella: www.vound-software.com

- Lexbe Online: www.lexbe.com
- Logikcull: www.logikcull.com
- Nextpoint: www.nextpoint.com
- Quick View Plus: www.avantstar.com/products
- Z-Discovery: www.zapproved.com
- Wind: www.windlegal.com

You can also keep costs under control by narrowing down the data to review first. Focus on what you know about what you are seeking. Think about specifics such as the date range and key players. This will allow you to narrow the data down to a smaller set of documents through metadata analysis. This doesn't mean you should completely discard other data, but it can help you to save time and money if you locate what you need early on.

Falling Into the 'Scope Creep Trap'

The flip side of needing to make sure you don't miss any important information or data sources and looking individually at unstructured files is that it is all too easy for the scope of your search to become too broad. You believe there may be a single text that can help your case, so you look through six months of text communication. When you come up empty, you decide to expand your search, looking back a year. Then two. Or you decide that you aren't looking for a text, but an email.

Many lawyers—even those who have been practicing for many years—do not focus enough on the e-discovery process. Perhaps they believe there are more important issues to deal with on a case. Or that the data they need to go through will be relatively small. Or that the process itself has become so automated that it will be a breeze.

"Project parameters" for e-discovery are inherently volatile as new facts are uncovered in a case and potentially useful data sources are discovered. Unfortunately, what this means is that it is all too easy for a seemingly simple e-discovery process to become bloated both in time and money spent.

Solution: Project changes are to be expected, but even small changes can add up over time. Monitor the situation regularly, and have key stakeholders sign off on proposed changes. Set a budget at the start of e-discovery, and make sure all scope changes take this factor into consideration.

What is the likelihood of finding the information you need? Why is the scope changing—simply because you didn't find what you need? Or is it because new information has come to light that leads you to believe you were looking in the wrong place? Can those discovery dollars be spent better elsewhere?

Not Starting Early Enough

Many lawyers—even those who have been practicing for many years—do not focus enough on the e-discovery process. Perhaps they believe

there are more important issues to deal with on a case. Or that the data they need to go through will be relatively small. Or that the process itself has become so automated that it will be a breeze.

Whatever the reason, they wait to begin the e-discovery process. This puts them behind the eight ball when they need to complete e-discovery quickly. They may end up with incomplete data, information that is incorrect, or a court violation if they are unable to comply with a data request.

Solution: The best way to ensure that e-discovery goes as smoothly as possible is to start the process immediately. Do your best to anticipate requests and determine where the more valuable information is likely to exist.

E-Discovery is Done by Project

While it is true that every case and every matter is different, that doesn't mean your e-discovery practices need to be reinvented each time you take on something new. Sadly, though, this is exactly what happens in many instances.

An ad hoc, project-based approach to e-discovery is taken because legal professionals often have to find unique ways to track down information they believe will be pertinent and helpful. But constantly having to figure out how to conduct e-discovery in a particular case leads to a number of problems.

First, it adds time and money because you have to come up with the process on the fly rather than diving in and actually beginning e-discovery immediately. Second, it can make clients feel like you are milking them since you don't have a specific process in place before you begin—why should they pay for you to figure it out?

Solution: To fix this problem, law firms need to work with businesses to enact better information management practices that use comprehensive technology solutions to increase data organization. This will not only speed up e-discovery, but set down a clear plan that can be followed for these types of cases in the future.

Your Analytics Just Aren't Good Enough

Of course, most law firms don't want to handle e-discovery on a case-by-case basis. You want there to be a clear process, because it will make you more efficient and clients will be happier.

But how do you come up with that process? By utilizing more useful metrics.

Solution: You need to use software that provides you with spending totals per matter, spending totals per file type, spending

totals per location, volume analytics across cases, production statistics (including how many times a document has been produced), collection and preservation information that shows past discovery activity, and search metrics.

These things can help you to estimate the necessary time and budget, come up with an overall e-discovery process, and generally streamline things for your clients—and your firm!

A Lack of Convergence

Chances are good that you have digital e-discovery tools. In fact, you may have several of them. Maybe you invested in a review tool as well as an Early Data Analyzer. And then something else to handle file processing. And so on.

The point is that you opted for these tools to streamline the process to make it faster and more affordable, but what you've actually done is lost a lot of time and money. Every time you move from one software “solution” to the next, all of your electronically stored information has to be exported and processed by the tool down the line. And that's assuming that the new and old softwares don't have any communication issues that can slow things down or even cause information to be lost entirely.

Solution: In the last few years, technology in this area has evolved significantly. Law firms started to recognize that a single tool that was able to handle e-discovery from beginning to end would be much more efficient, and legal service providers listened. While there is certainly plenty of room for improvement, there are a number of tools and services that handle multiple aspects of e-discovery without having to continually move data around.

Some of the best out there include Logikcull, Nextpoint, CloudNine, and Wind.

Unwilling to Work With the Opposition

Your job as an attorney is to win cases, which means that the other side has to lose, right? They are the enemy, and you should oppose them at all costs. But where e-discovery is concerned, is that really the way to go?

One of the biggest issues that drives up costs for e-discovery is one side sticking to unreasonable exchange protocols. This can literally price smaller firms out of e-discovery by demanding they use services that far exceed their budget.

Solution: If you run into this kind of situation, you have two choices—battle it out in front of the judge or play nice with the other side in an effort to get them to agree to use lower-cost e-discovery solutions.

Difficulty Recovering E-Discovery Costs

The laws in place for recovering e-discovery costs are slowly evolving. However, they are still largely, for lack of a better term, “analog.” For example, they refer to the price of making copies or the cost of getting to data that is difficult to access.

But unfortunately, access isn't the problem that legal professionals face in our increasingly digital world. Getting data is easy, generally speaking. What is currently driving up the cost of e-discovery is the fact that there is so much of it out there that it takes an inordinate amount of time and money to review. There have been anecdotal accounts of \$4 million lawsuits where \$3 million was spent on e-discovery—\$3 million that is not recoverable.⁴

Solution: Sadly, this isn't something that individual lawyers or

firms can fix, but rather a systemic problem. Courts are starting to get more creative in reining in disproportionate e-discovery, but it's still a big issue.

Laws Are Complex and Ever-Changing

E-discovery laws need to evolve as new technologies emerge and big data becomes even bigger data. But when legal cases involve international elements, trying to navigate the disparate legal systems across the globe and keep up with the hundreds of laws added each year can be incredibly daunting. What was true six months ago may no longer apply.

Much in the same way that e-discovery used to largely impact big mass torts and civil-action lawsuits, but now affects smaller cases as well, international law is coming into play in more and more cases. Data can now be easily housed anywhere in the world, which sometimes brings up questions of who owns it and what laws apply.

Solution: There's not much to do here other than strive to keep up with laws related to data ownership and e-discovery as much as possible. Perhaps one day there will be unified international data laws in place, but until then, it is up to legal professionals to navigate this tricky space to the best of their ability.

Comparing E-Discovery Providers Is Next-To-Impossible.

Even while more cost-effective e-discovery tools have become available and providers of e-discovery services have lowered their prices, it is still incredibly difficult to price compare. Simply put, the tools, services, and price points differ from solution to solution—but they do so in a manner that makes it almost impossible to tell how much you might end up paying for e-discovery services using one provider or tool over another.

Solution: More standardization is needed to alleviate this problem, and possibly more regulation and oversight. But until that becomes a reality, your best bet is to work with a legal IT company that can advise you on what software solution best fits your case or firm. They'll be able to draw on the experience of all of their clients to advise you for your specific situation.

Technological Incompetence

A lot of lawyers probably won't want to hear this, but a huge part of the problem is that, as a group, they're just not as tech savvy as they need to be. This causes all kinds of problems with e-discovery, including overpreservation of data, overpayment for services, methods that are outmoded, shortsighted source identification, using outdated tools, failing to identify and address new issues, and discovery that is poorly executed and unfocused.

In particular, experts point to a lack of understanding about the nature and pervasiveness of smartphones, BYOD (bring your own device), texting, and instant messaging versus email, as well as social media in general. It's not that lawyers do not know about these things, but that their understanding does not go deep enough to allow them to adequately navigate and explore them as it pertains to e-discovery.

Solution: Don't go it alone. Hire the right project manager or project management company. Beyond experience with e-discovery, you want to seek out someone who can share their expertise effectively. The goal shouldn't be just for them to manage the project, but also to guide and educate your legal professionals through-

out the process. You'll have better, more cost-effective results if everyone is informed and working as a team.

Data is Sorely Mismanaged

Rare is the company these days that truly has a good handle on its data and knows exactly what it needs to hang on to—and where that data lives. Mismanaged data is bad for e-discovery because it is more difficult for you to track down and—since there is likely to be a lot of extraneous information—it takes longer to sift through to find facts and details pertinent to the case.

Solution: Unfortunately, this is an issue that starts with businesses and individuals themselves. If you are a lawyer taking on a case, you have to step in and deal with the situation as it is. In-house attorneys for corporations, however, should pay careful attention to data management and suggest comprehensive solutions regarding how to deal with it, including creating an information governance policy that can allow companies to defensibly delete data.

Conclusion

If you wanted to sum up the problems with e-discovery, it would be this: The process is completely and utterly erratic. Costs vary. Laws vary. Processes are underdeveloped. And data keeps growing and spreading to more places. As a society, we all need to get better at managing all of the information and data in our lives.

But since attorneys often have little say in how data is managed before they take on a case, it is incredibly important to pay attention to emerging trends in communication technology and business and utilize innovative solutions to increase efficiency. ☺



Andres Hernandez is the co-founder of Wingman Legal Tech, which is a technology consulting firm that specializes in law firm technology. Their technology solutions are designed for simplicity and efficiency.

Endnotes

¹*The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things*, Infographic, Cylcone Interactive (2014), <http://www.emc.com/infographics/digital-universe-2014.htm> (last visited June 14, 2016).

²Case Study: Predict Proves Effective for Small Collection, Catalyst Repository Systems (2015), available at https://www.catalystsecure.com/pdfs/case_studies/Catalyst_CaseStudy_Predict_Proves_Effective_for_Small_Collection.pdf.

³Tom O'Connor, *Cost-Effective E-Discovery for Small Cases*, A.B.A. (2013), http://www.americanbar.org/publications/gp_solo/2013/january_february/cost_effective_ediscovery_small_cases.html (last visited June 14, 2016).

⁴Steven Baicker-McKee, *The Award of E-Discovery Costs to the Prevailing Party: An Analog Solution in a Digital World*, 63 CLEV. ST. L. REV. 397 (2015), available at <http://engagedscholarship.csuohio.edu/cgi/viewcontent.cgi?article=3846&context=clevstlrev>.

**Missing Your
Target Market?
Hit the Bullseye
with Our Reach**



Partner with the Federal Bar Association today—contact Heather Gaskins, Director of Development at (571) 481-9106 or hgaskins@fedbar.org to hit your target!