THE

COMPREHENSIVE GUIDE TO E-DISCOVERY DATA COLLECTION

2ND EDITION

Overcollection is one of the leading drivers of exorbitant e-discovery spending, which means data management and information governance play a vital role in allowing legal teams to find and collect the right data quickly and efficiently. By retaining data that is vital for business and legal operations, and organising that data in a way that anticipates legal needs, you can streamline your e-discovery process, giving your team a competitive edge.

By mastering e-discovery data collection best practices and lessons learned from experienced e-discovery practitioners, you can ensure your collection and processing processes are not only defensible, but also allow you to get to the facts of the case much sooner—saving time, money, and human resources.

exterro

THE GUIDE, IN BRIEF

In this e-book, you will learn:

- The fundamentals of e-discovery data collection
- Tools and techniques for efficient, defensible data collection
- Advice from experienced e-discovery practitioners

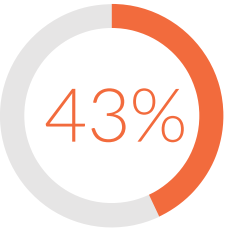
Contents

Key Data Points on E-Discovery Collection

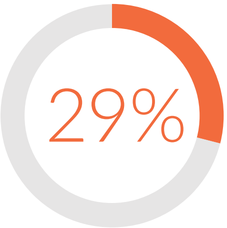
Exterro asked 208 in-house legal and IT professionals involved in E-Discovery:

"What is your biggest obstacle in locating potentially responsive data?"

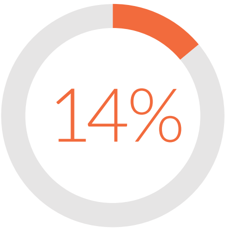
Findings from this survey indicate searching through vast amounts of electronically stored information (ESI) to find responsive data is the number one challenge for both IT and Legal teams at global enterprises. Identifying and accessing data sources for collection was reported to be the biggest obstacle by the second most number of respondents.



OF LEGAL DEPARTMENTS STILL "COLLECT TO PRESERVE," POTENTIALLY INCREASING E-DISCOVERY COST AND RISK



OF LEGAL DEPARTMENTS COLLECT MOBILE DATA FOR LITIGATION "ALL OR MOST" OF THE TIME



OF LEGAL DEPARTMENTS COLLECT SOCIAL MEDIA DATA FOR LITIGATION "ALL OR MOST" OF THE TIME

IN 2014

EACH IT PROFESSIONAL WAS RESPONSIBLE FOR **230** GB OF DATA

DATA

BY 2020

EACH IT PROFESSIONAL WILL BE RESPONSIBLE FOR **1231** GB OF DATA

SOURCES: *Exterro and ACEDS 2018 In-House Legal Benchmarking Report* | *Exterro and FTI Consulting The State of E-Discovery 2019*

SURVEY RESPONSES

Why is searching and collecting large amounts of data so challenging?

Attorney

"My organisation still has a large amount of data in hard copy form that is located in a number of offices throughout our service territory. There is a high risk of us not finding responsive documents simply because we don't know they exist."

IT

"We have a huge volume of unstructured data. Identifying and collecting from this data is daunting. The sheer volume also is a barrier to proper governance and retention."

Paralegal

"Data in our company is everywhere, so finding the right sources is a challenge, and sometimes it is difficult to get the data out of the sources they reside in."

CHAPTER 1
The Fundamentals of E-Discovery Data Collection

Data collection is perhaps the most technically rigorous and complex of all the e-discovery phases. It involves the extraction of potentially relevant electronically stored information (ESI) from its native source into a separate repository. Because collection involves direct interaction with data, most people mainly associate it as an IT activity. However, both legal and IT professionals must collaborate to develop the right collection strategy for an organisation.

Here are 3 things you need to know about e-discovery collections:

1 Primary Data Collection Challenges

- **Fragmented or Insufficient Tools:** Most legal teams don't have search/collection tools to help with the discovery process, and if they do, they often have to use a combination of tools, at times from different vendors.
- **Search Definition:** Knowledge of correct keywords and stopwords, as well as indexing various forms of ESI—such as email attachments, embedded audio, and metadata—are all required for a successful search.
- **Undercollection Creates Risk:** Traditional search methods that rely solely on keyword searches can often miss responsive data, which can result in counsel developing an incomplete or inappropriate case strategy.
- **Overcollection Increases Cost:** Overcollection of data increases the cost of already expensive activities on the right-hand side of the EDRM, including hosting fees and document review.

2 Data Collection is *Not* Preservation

Even experienced legal professionals tend to conflate preservation and collection. While collecting as a way to preserve certainly would meet the court's intention, it is a very costly and inefficient way to do so. Think of preservation in terms of ensuring potentially relevant data isn't deleted. Courts don't prescribe a particular method for preservation, they just require that it gets done. Collection, on the other hand, is the first tangible step towards producing documents to the other side. While certainly not all collected documents will ultimately get produced, the idea is that collection feeds into the review process, which in turn dictates the production set.

But don't assume collection involves collecting all data that is preserved.

Defensible preservation practices don't require counsel to collect everything from every potential custodian that might have responsive ESI even when a claim is filed. A reasonable practice is to interview candidate custodians and "tier" them according to how likely their ESI will be implicated in the matter. Those at the top will likely merit having their email and home directory files collected; those further down may just receive a legal hold notice to preserve ESI related to the matter until further notice. Custodian interviews also have the benefit of identifying additional custodians or data sources that counsel didn't consider previously.

A newer, more efficient option involves using preserve-in-place tools that prevent certain files or folders from being deleted until the hold administrator obtains approval. We will talk about these collection approaches later in the guide.

3 The Role Data Processing Plays in the Collection Process

Rather than cover data processing in its own section of the guide, we've decided to include it with collection, since the two are so closely intertwined. Processing prepares collected data for attorney review. After collection, the resulting document set will include a rather messy mix of file types and formats, attachments, meaningless system files, and plenty of duplicates. Processing cleans up the mess and formats the collected ESI so that it can be culled and searched by attorneys and review tools.

We won't get to into the weeds on data processing, since it's a highly technical process that includes a lot of concepts and jargon that the average e-discovery practitioner doesn't need to know. What is worth discussing, however, is who actually does the processing. Traditionally, most organisations outsourced data processing to third-party vendors who would use specialised technologies to winnow data sets down and deliver them back to clients for next steps. Today, many companies still outsource processing, but there are a growing number of companies who use in-house processing software. There is also an emerging class of tools with "one-click collection" abilities that consolidates collection and processing into one step. There will be more on these tools later in this guide.

CHAPTER 2
ESI Types: What Must Be Collected

Defensible e-discovery practices must account for virtually every form of ESI. And while it's one thing to identify and preserve various forms of ESI, it's often quite another to actually collect it. Different data sources have different levels of accessibility and present different collection challenges. Here is a breakdown of six common categories of ESI that you might need to collect for e-discovery.

1. **Most Active:** Data that you interact with on a regular basis, such as email and other traditional files that are stored on a local hard drive or network drive. This ESI tends to be fairly easy to access and collect.

2. **Cloud:** The fastest growing category of ESI is data stored on cloud servers (e.g., SaaS applications, cloud storage, social media), including Microsoft Office 365®. Cloud providers may have different policies around accessing data for collection, so you should familiarise yourself with these details before you actually need to collect.

3. **Mobile:** Mobile devices are so ubiquitous that organisations should be prepared to collect from both company and BYOD devices, but often they are not. Mobile collections require sophisticated tools and expertise, so legal teams must have a plan in-place to collect mobile data, especially from key custodians.

4. **Offline:** Offline data that is no longer in active use but is stored or archived. Even though offline data can't be accessed over a shared server, collecting it usually presents fairly minimal challenges as long as you know the physical location of the data and the system on which it's stored.

5. **Backups:** Traditional backup tapes or disaster recovery systems are designed to store data in the event that it must be restored. These systems compress files and are not easily searchable or accessible, and therefore they tend to present significant collection hurdles.

6. **Hidden:** Previously deleted or fragmented files that exist on various systems and are usually not readily visible to regular system users. These files are highly inaccessible, and attempting to recover them requires specialised tools. Learn more on hidden files in our section on forensic imaging.

CHAPTER 3
Approaches for Data Collection

There are a variety of ways that organisations approach the collection process. Questions that might dictate the collection methodology might include:

1. How much data is involved in the legal matter?
2. How many sources of data are implicated, and how accessible are those data sources?
3. Will the collection involve any specialised tools or expertise?
4. Does the legal matter involve encrypted or sensitive data?
5. Are there internal IT resources available to perform/assist with the data collection?
6. What are the time constraints (production deadlines, retention schedules, etc.)?
7. What type(s) of collection technologies are deployed to perform the collections?
8. Is the case civil or criminal?

Answers to these questions will help determine which of the following collection approaches to employ.

THE RISKIEST
Employee Self-Collection

Probably the riskiest of all collection approaches, employee self-collection involves letting the custodians themselves copy relevant files into a shared drive or portable storage device. Most experts advise against employee self-collection, pointing out that most employees aren't technically savvy and are highly likely to make mistakes and overlook key documents. Likewise, several courts have also questioned whether employee self-collection constitutes a "defensible" e-discovery response. That being said, for small matters involving low volumes of highly conventional data (email, word processing documents, etc.) employee self-collection may be reasonable and cost effective, especially when the opposing party and judge have signed off on the plan ahead of time.

THE MOST COMMON
IT Collection

By far the most common collection approach, IT collection involves members of your IT department performing the actual data collection at the direction of the legal department. On the surface, involving IT professionals in the collection process makes sense, since they understand the data landscape and usually possess the technical skill to get everything that's needed. However, there are downsides. In organisations with limited internal IT resources, data collections can be time consuming and keep IT professionals from other business-critical projects. As mentioned above, IT professionals also tend to associate data collection with forensic imaging, and without clear guidance from the legal team on what specifically to go after, they are likely to collect very broadly, resulting in more data that has to be processed and reviewed, driving up e-discovery costs.

THE 3RD PARTY WAY

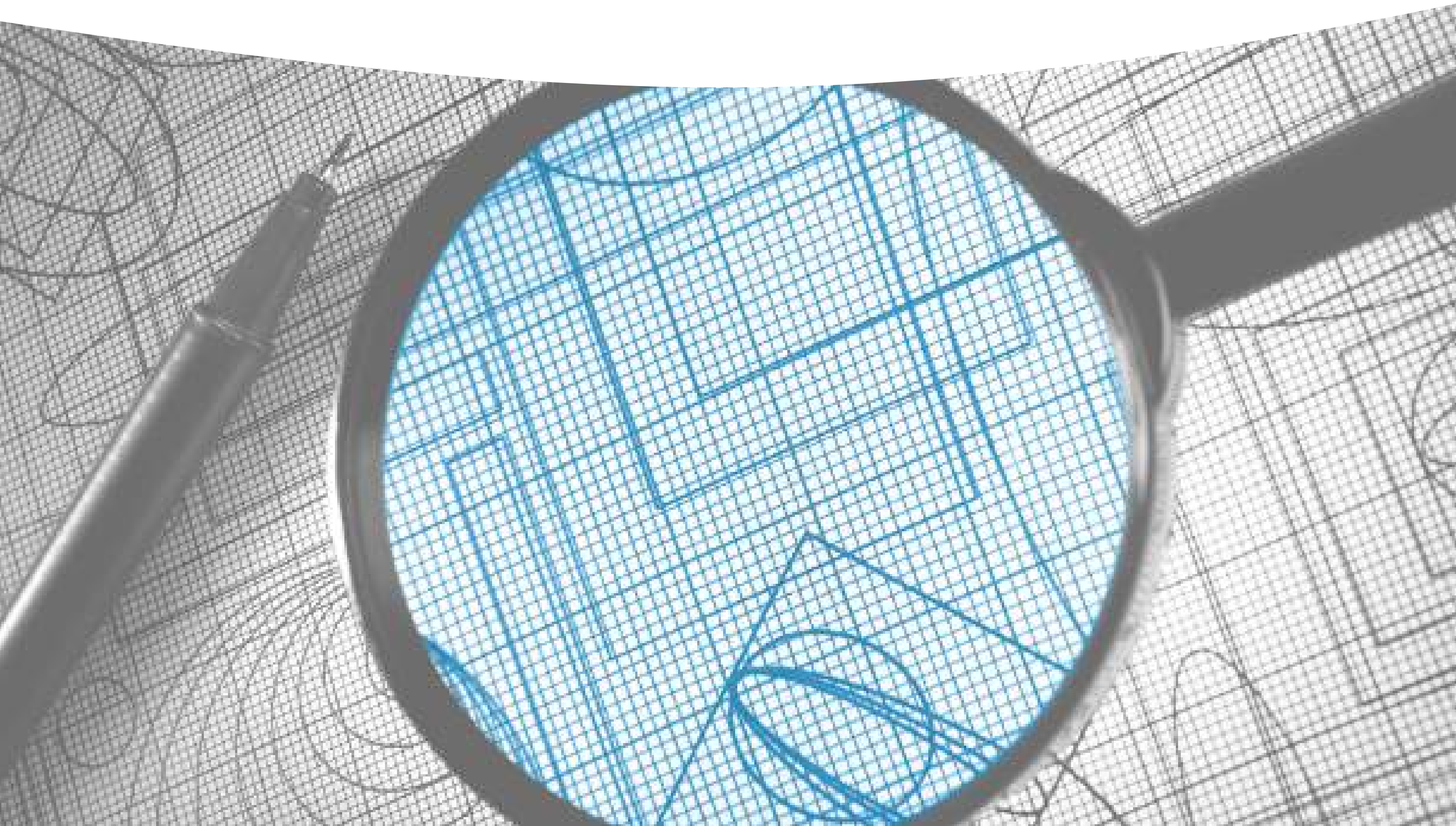
External Collection

For organisations with very limited IT resources, a third party expert might be called on to perform the data collection. An outside expert is likely to have set procedures and all the necessary tools and skill to perform a collection that will withstand the highest levels of judicial scrutiny. But there can be a considerable expense associated with bringing in outside assistance, which is why most experts advise that organisations make it a priority to establish at least some level of internal collection capacity.

THE STREAMLINED APPROACH

Remote Collection

These collections employ a centralised internal collection system that is integrated with company data sources allowing for ESI to be collected remotely. Though the collection might still be performed by an IT professional, it doesn't require any direct interaction with the data sources themselves and can usually be performed more efficiently than traditional methods. These systems also support more targeted collections by applying search and analytics technologies. While this type of collection technology does require an upfront investment and deployment process, it's the most efficient and cost effective approach for large organisations with consistent collection demands.



CHAPTER 3 DEEPER DIVE

Important Technical Considerations for your Data Collection Approach

We live in a digital world, which means that regardless of your profession, you need to have some tech savvy. This couldn't be more true for legal professionals. Long gone are the days of paper files, and with growing electronic data volumes, as well as constantly changing data types, understanding the technical nuances involved in data collection is a must.

Role of Metadata in Collections

You can't address collection in e-discovery without talking metadata. You'll come across different definitions for metadata (like all things e-discovery, it seems one definitive explanation for a concept, process, or activity is never enough). Our favorite definition is that metadata is the data about the data. Let us explain:

When you look at regular document on your computer you see the words in the document of course, along with the name of the file, and where it's located. Information about when a file was created, modified, last updated, who made edits to the document, etc. may not be terribly useful on a day-to-day basis. In the land of e-discovery, this contextual information can be hugely important and has to be included when the document is ultimately collected.

If you're interested in technical detail, read more about MD5 hashing, one way of understanding if files are identical or have slight differences.

Knowing the Difference Between Forensic Image vs. Logical Copy

If you tell an IT professional that you need to collect data from a computer hard drive, chances are you are going to be presented with a forensic image of the drive (also known as a "bit by bit" or bit stream copy). At the most basic level, a forensic image is a complete copy of a drive—including the portions of the drive that aren't allocated to active files (known as slack space). It is what would normally be considered an exact duplicate. These types of images give you both the files you'd expect to see if you were browsing a file listing, and also data from previously deleted files. Forensic imaging requires specific tools and is usually administered by an expert.

Alternatively, a logical copy is simply a copy of the contents of the directories on a disk and does not include previously deleted data or other information that a forensic image would capture. They are also much less technically intensive and can be performed by just about anyone with a little training and the right software.

So Which Is The Best Approach For E-Discovery?

Most experts will tell you that in almost all civil matters, a logical copy will meet the court's expectation. There is certainly a place for forensic imaging, but it's usually only necessary when there is a suspicion of data tampering or in cases where previously deleted files are at the center of the controversy.

CHAPTER 4
Tools and Software for Data Collection

Data collection is not a one-size-fits-all endeavor. There are a variety of tools and capabilities that you can deploy depending on your specific collection needs and priorities.

Here are some specific e-discovery collection capabilities you may want to consider.

Concurrent Processing

Processing has traditionally taken place post-collection as a separate activity, typically handled by service providers who charge on a per-gigabyte basis. However, new search and collection technologies process data at the point of collection, eliminating the need to send collected data to a third-party vendor.

Pre-Collection Analytics

Pre-collection analytics have a huge influence on the collection process. These tools crawl data sources and deliver insights into document volumes and can also perform more advanced searching and filtering. They equip you with the necessary intelligence to and focus on relevant content.

Single-Click Collection

Leveraging the insights gained from pre-collection analytics, single-click collection allows e-discovery professionals to collect only the documents identified as relevant during early case analysis. This highly targeted collection helps teams to the facts of a matter quickly and reduces downstream costs, like data hosting fees and document review.

Data Source Integrations

Integrating your collection software with your enterprise data sources (email servers, Sharepoint servers, structured databases, etc. can greatly streamline the collection process by eliminating the need for IT to conduct manual collections. Integrations allow remote collection, and they also minimise the technical complexities surrounding collections, allowing non-IT professionals to be more involved in the process.

Spot Collectors

Even when you have an integrated collection environment and can collect over the network, there are still instances when you may need to grab data off a system that isn't connected to the network, such as a remote employee's laptop. Spot collector tools are portable USB devices that allow IT professionals or custodians to crawl and collect off non-network systems. These tools can be pre-configured to collect only relevant files rather than complete copies of a computer's hard drive.

Mobile Collection Tools

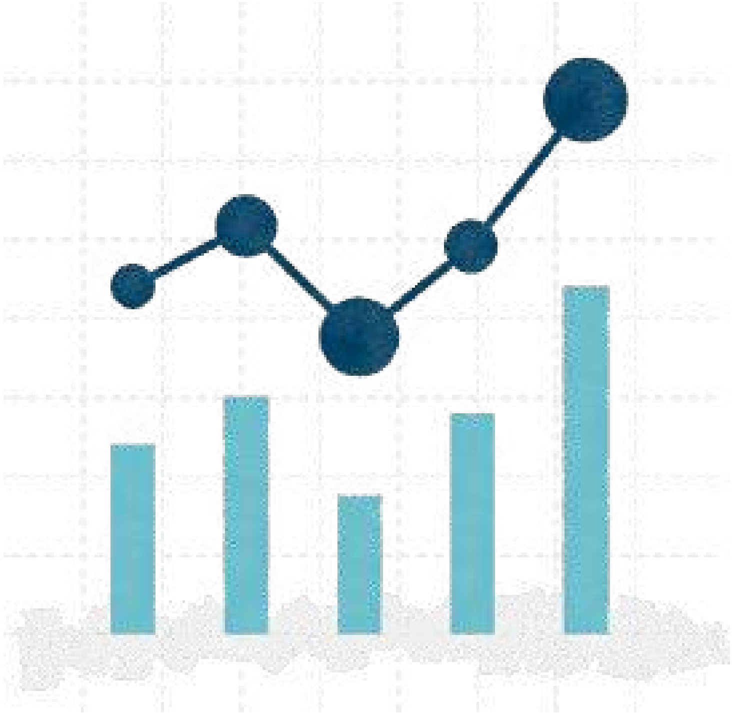
We discussed the challenges of collecting mobile data earlier. Ideally, any data on a mobile device will be located somewhere else that is a little more accessible, such as an email server. But workers in some industries create content that never leaves their phones—like text messages—that may need to be collected. Fortunately, there are specific devices that are designed to extract data off of mobile devices and reformat it for the purposes of attorney review and legal production.

CHAPTER 4 DEEPER DIVE

New Technology to Minimise Collection Volumes – In-Place Early Case Assessment

Understanding ESI Before Collection

In-place early case assessment (ECA) leverages increased visibility into ESI, **before you have collected it**, to define case strategy. It allows e-discovery professionals to get to the facts of a given matter faster, producing downstream cost and time savings during collection, review, and production.



Using a broad set of analytic and predictive intelligence capabilities, in-place ECA rapidly identifies the most important documents prior to collecting a single document, justifying proportional and narrowly tailored e-discovery production parameters. Bob Haskin, Managing Director at Morae Global, explains the power of true early case assessment, "It shifts from traditionally reactive approaches to a more proactive one with the goal of learning what you need to know sooner in the process."

Using more advanced technology, such as artificial intelligence, during ECA, can provide e-discovery professionals with even more insight, revealing hidden concept clusters, communication patterns, and custodian relationships. Tools like Exterro Smart ECA can unlock even greater savings than traditional ECA technologies, speeding up organisations' ability to get to the facts and define case strategy.

Implementing In-Place Early Case Assessment

With In-Place ECA technology, legal teams leverage document storage and preservation capabilities to access data earlier in the EDRM. Since these capabilities are integrated with commonly collected data sources, legal teams are empowered to quickly review data in-place, before collection.

Linda Luperchio, Information Governance and E-Discovery Director at The Hanover Insurance Group, recalls, “Because I’m not waiting for everything to be collected and processed, I can just start looking at the data right away. Usually within 48 hours I can provide a hit report.”

> “More and more technology providers are working to embed native ECA functionality directly within applications, as opposed to companies needing to use third-party tools. If many of your e-discovery data sources offer ECA functionality, the return on investment from a third-party in-place ECA tool might be limited.”
>
> *-Bob Haskin, Managing Director at Morae Global*

To provide effective insight, in-place ECA technology requires:

- → Integration with data sources
- → An accurate data map
- → Appropriate permissions for sub-folders
- → Search functions across content keywords and metadata

Alternatives to In-Place ECA

There are two primary alternatives to dedicated in-place early case assessment technology: traditional e-discovery workflows and ECA capabilities that are baked into other software applications.

The biggest competitor to in-place early case assessment isn’t another technology; it’s entrenched ideas about e-discovery workflows. When the EDRM was developed, technology simply didn’t allow deep insight into ESI prior to collection. Attorneys could negotiate e-discovery protocols using boilerplate arguments, because they could safely assume their counterparts would be in the same situation. In-place ECA upends that paradigm.

On the technology front, many data repositories provide some level of analytic insight into their contents. These native capabilities also compete with third-party in-place ECA technologies.

CHAPTER 5
Collection Strategy: Developing Your Own

Your collection strategy will change with every matter. In some cases—for example, very high stakes legal matters involving precarious data sources - it may be wise to collect data immediately. In other matters, immediate collections may not be necessary, especially if you have a strong preservation process in place. It's common for litigants to collect very highly relevant data early, since they know it will need to be collected eventually, but collecting very broadly in the early days of a matter is usually not advisable, as this will typically just drive up your costs with very little associated benefit to your case.

It's also important to consider how your case strategy impacts your collection strategy. If your case is inevitably headed for an early settlement, it probably doesn't make a lot of sense to collect and process a bunch of data that ultimately won't be needed. Other considerations that should go into your collection strategy include whether outside experts should be involved, if there is any sensitive data that warrants greater protection measures, and whether any employees—like a person named in an incriminating lawsuit —might have incentive to alter or delete relevant data, in which case a more proactive collection might be warranted. On top of these considerations, legal teams must now evaluate how their collection aligns with new FRCP rule 26, ensuring your collection approach is proportional to the matter at hand.

Collection Strategy Begins with Preservation

Like any good strategy, planning begins many steps ahead of the actual execution of the task. In this case, a good collection strategy actually begins with the preservation stage. Having a forward-thinking preservation process that anticipates collection can give you a competitive edge by allowing you to get the facts of the case sooner and adjust your strategy accordingly.

Here are 5 Things to Consider for Collections during the Preservation Stage:

1. Use Early Case Assessments to Refine Your Case Strategy

Early Case Assessment (ECA is becoming a must-have for legal teams based on the FRCP's new emphasis on proportionality. With ECA software tools, legal teams can get an estimate of the amount of data likely to be responsive, which is vital information to have when walking into Rule 26(f conferences. So, what is ECA?

Early Case Assessment is identifying how much potentially relevant data is related to your case by using a sample search of key custodians/ data sources, and analysing it for determining the relevance criteria (file types, search terms, relevant custodians, date ranges, etc.. These metrics can be used to develop a budget and to prepare your team to negotiate and fight for favorable e-discovery parameters at your Rule 26(f conference.

2. Leverage In-Place Searching Before Collection

This is one of the least known techniques being used today, but it can be a powerful weapon with incredible cost savings. Traditionally in ECA, you must collect a sample of the client's data, but now, some of the e-discovery solutions available can actually go out and look at the data before collection. You can view documents and the relevance of search terms, allowing you to figure out the relevance criteria, prepare for the Rule 26f conference, and develop a budget without collecting data. It can also be leveraged in the context of when you don't know much about your case and are conducting an investigation prior to litigation, or even just a pure internal investigation. Collecting data is an expensive endeavor—if you have the ability to index data without having to process it and collect it, this is a real game changer.

3. Negotiate for Proportional Collection Requirements

Once you've preserved everything you need to preserve, then you can engage in negotiations with the opposing side. Since the updated FRCP took effect, this process takes place much sooner than in the past. By negotiating about what truly is discoverable, you can then define and give guidance about what you are going to collect. You can also use technology to determine the scope of discovery before the Rule 26f conference, which can then help case strategy regarding making a persuasive and fact-based proportionality argument.

4. Create an Audit Trail

It's vital that you document your e-discovery processes, because you may have to defend them. There are lots of moving parts in e-discovery, and it's inevitable—given the complexity and the amount of information—that you may drop the ball on something. If you can document your processes and demonstrate to a court that you were reasonable, even though you may have lost something, chances of sanctions are small. If you have powerful technology solutions on your side, then that process can be automatically captured within the e-discovery system. Otherwise you have to go the old route with spreadsheets or even paper files.

5. Don't Over-Engineer Your Collections

In e-discovery you are going to have to collect data, and usually there are no ifs, ands, or buts about it. But the mere fact that you have to do it doesn't mean you have to overthink how you collect. Remember, for civil litigation, and when a likelihood of fraud or bad faith activity (e.g. disgruntled employee, etc.) isn't apparent, bit-by-bit, forensic collections should not be conducted. They are expensive and collect everything on a data source, including computer generated files and other irrelevant data. To be defensible, data collections should include the collecting of a document's metadata, and that a chain of custody is maintained (i.e. MD5 hashing).

Relevant Case Law:

New Mexico Oncology and Hematology Consultants, Ltd.
v. Presbyterian Healthcare Services

SOMETIMES JUST SENDING A LEGAL HOLD IS ENOUGH

CASE ENTRY

New Mexico Oncology and Hematology Consultants, Ltd. v. Presbyterian Healthcare Services (D. New Mexico Aug. 16, 2017)

WHY THIS CASE IS IMPORTANT

This case re-iterates the fact that a collect-everything approach is most likely not needed. Just ensure your preservation practices are "reasonable" under the circumstances, which could mean sending a detailed legal hold notice to key custodians.

Overview

In this lengthy antitrust and RICO dispute between healthcare organisations, the plaintiff moved for spoliation sanctions based on the defendants' legal hold procedures.

The plaintiff had four qualms regarding the defendant's preservation process:

(1) Too much employee discretion given to determine relevancy;

(2) Employees were forced to delete or archive potentially relevant emails based on their data retention policies;

(3) Not enough employees were initially placed on hold (35 initially, which was expanded to 209 total);

(4) A server-side hold was put into place for all key custodians.

Ruling

1. **Rejected Motion Due to Lack of Prejudice.**
   The court rejected most of the plaintiff's arguments due to the lack of evidence to prove prejudice or bad faith.

2. **Reasonableness, Not Perfection Required.**
   Upon review, the court found that the plaintiff had a reasonable preservation process, including a detailed legal hold notice, which advised "when in doubt, preserve" and instructed the first custodians on hold to refer other key custodians.

3. **Implement a Server-Side Hold.**
   Again, without evidence that the plaintiff was prejudiced by the defendant for not implementing a server-side hold, the court could not find sanctions. The court did however state "the best approach is to implement a server-side hold on all digital data utilised by key employees and to later use search algorithms to parse relevance."

HOW TO ENSURE YOUR COLLECTION IS DEFENSIBLE

Collection can lead to a lot of contentious disputes between parties when there is suspicion that not all relevant data was collected, or that the collection process itself altered the contents of the data. When such controversies surface, parties typically rely on a few mechanisms for proving that a collection was conducted in a defensible manner. These include:

CHAIN OF CUSTODY

E-Discovery think tank The Sedona Conference defines chain of custody as the "documentation regarding the possession, movement, handling, and location of evidence from the time it is identified to the time it is presented in court or otherwise transferred or submitted." A thorough chain of custody log is designed to demonstrate the authenticity of a document and disprove any claims of data tampering.

MD5 HASHING

Commonly referred to as a "digital fingerprint," a hash value is a special encryption code that is associated with each computer file. The purpose of a hash code is to provide files with a unique identifier. If a file's contents or metadata change, the file's hashtag will change as well, indicating that the file is not the same as it was before. By comparing hash values before and after collection, you can easily show that a file is the same pre-collection as it is after. For more on hashing, read e-discovery attorney Ralph Losey's terrific blog post on the topic.

AUDIT TRAIL

Audit trails are automated records generated by systems that track user activity. In the context of collection, they can be helpful in showing when a collection took place, what it entailed (collected data amount), and which user initiated it if such information is ever requested by a judge or adversary.

CHAPTER 6
Final Key Takeaways: 4 Must-Follow Best Practices

Data collection is a dynamic and multi-faceted process that relies on sound e-discovery strategy, as well as solid technical resources and expertise. There are important best practices that fall under each of the various elements of the collection process, but here are four big ones you should know.

BEST PRACTICE #1

Don't Over-Collect, Target Your Collections

We know that it's easy to identify a relevant custodian and copy his or her entire hard drive or email folders. But easy doesn't equate to smart. More data collected means more data processed and ultimately reviewed. And that all adds up to more *money* spent on e-discovery. Instead, develop strong **preservation** and **early case assessment** processes, and most importantly target your collections so that you are only collecting the potentially relevant ESI, nothing more and nothing less.

Here are two central components of a targeted collection strategy that is both defensible and cost effective:

1. **Identify Key Players:** Virtually all litigation events can be traced back to a relatively small number of key custodians whose data will unquestionably be at the heart of the key issues of the matter. These individuals should be the early focus of any collection efforts, because their data represents the greatest risk. Starting small and building the collection plan out from the key players is a sensible way to corral spending.

2. **Talk to Custodians:** Key custodians not only have the relevant data, they also have information that can be extremely valuable in tracking down other responsive ESI and developing a case strategy. You should question key players about whether there are any other employees that may have been overlooked in the initial assessment of the case, as well as the nature and location of the data involved. Does the case mostly implicate email? Where did the custodian save his or her files? Were there any non-traditional forms of electronic data that may be relevant? Besides pointing you towards the relevant custodians and data sources, key custodians can also be instrumental in helping you develop a list of search terms for when it comes time to assess the data and start putting together the overall story.

BEST PRACTICE #2

Be Proactive

It's inevitable that some matters are going to present unique collection challenges. Maybe it's a case involving mobile data or one involving highly unorganised data on legacy systems. Whatever the case may be, do yourself a favor and recognise these challenges early on rather than right at the point where data needs to be collected. It's always better—and much cheaper—to assess your needs proactively to determine if outside resources will be needed and, if so, which vendors. Even if outside help isn't needed, it's important to give your internal IT team a heads-up that a potentially big project may be coming their way soon, so they can plan accordingly.

BEST PRACTICE #3

Integrate Your Collection Software and Data Sources

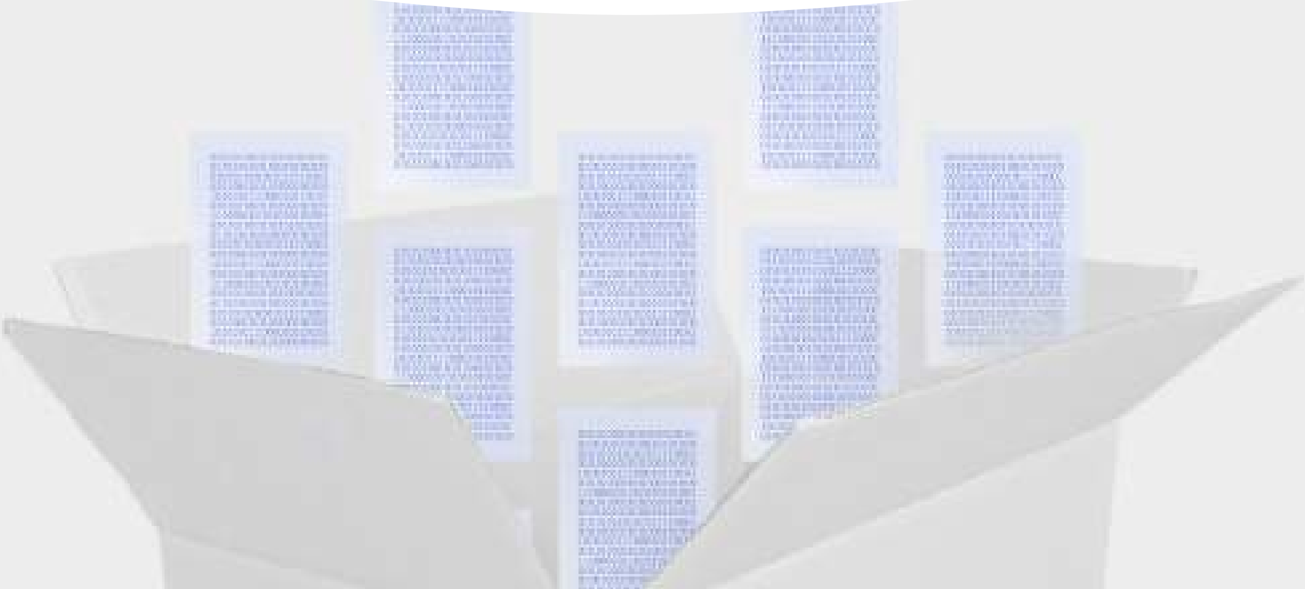
A lot of the best practices provided above have been process-based. On the technology side, the whole collection process can be greatly streamlined by integrating your collection tool with your corporate data sources. This allows collections to be much more targeted than traditional collection methods, like imaging an entire hard drive. You can also perform collections a lot quicker and limit disruptions to business processes that might depend on the targeted data repository. Advanced collection tools also have the ability to automatically produce reports (see chapter four) which also saves time and mitigates the risk of human error that comes with manual reporting.

BEST PRACTICE #4

Tier Your Collections

This relates to our point above about over-collection. It's always best to think of collection in terms of phases or tiers, rather than to try and do everything all at once. A tiered collection strategy involves prioritising data so that only the most highly relevant data is collected immediately, and less relevant data is collected only when absolutely needed. Just remember, that the only way to execute a defensibly tiered collection strategy is to have a very strong preservation process which gives you the ability to not collect everything immediately.

Some might worry about the defensibility of tiering and the potential loss of responsive data. To put these concerns at ease, tiering ensures both reasonableness and proportionality under the new amendments to the FRCP, because it relates directly to collection, not preservation. If you preserve broadly via legal hold, along with suspending your document retention policy, then you don't have to collect everything at the beginning of the matter. What's more, Judge Andrew J. Peck, Southern District of New York, has stated that the goal of these amendments is to avoid a collection of all the data, and instead focus on a collection of the data that is needed for the matter at hand. So, if you have a well thought out and documented process of what data was targeted and how the custodians were selected, that is the definition of reasonableness.



Conclusion

In many ways, e-discovery has changed drastically over the last decade, but in the end, it’s still about finding the electronic data that is relevant to a specific matter, and then meeting the requirement to produce that data.

What has changed and continues to grow more and more complex, is where that data comes from and how we share it—from electronic information, then to email, now to mobile, social media, instant messaging, and cloud based platforms—and many would say that e-discovery is now actually a subset of information governance.

The most efficient legal teams streamline their e-discovery process to collect the most relevant data quickly and efficiently. But this is it not something that can happen in an ad hoc fashion, but requires a mature process coupled with the right technology.

CHECKLIST
STEPS TO HELP CREATE AN EFFECTIVE, PROPORTIONAL COLLECTION PROCESS

○ Define criteria to determine the appropriate collection method (logical copy, forensic mage, etc.)

○ Define collection roles and responsibilities for legal and IT personnel

○ Develop standardised process for submitting collection requests from legal to IT

○ Determine capacity to perform internal collections focusing on collection type, data sources, file types and data volumes

○ Identify and document procedures for engaging third-party resources to perform collection/procedures

○ Establish preferred vendor list

○ Utilise targeted collection capabilities (date ranges, keywords, etc.)

○ Develop procedures to document the chain of custody during the collection process

○ Develop collection tracking process between IT and legal

○ Create standard processing specifications to ensure collection/ processing consistency

○ Develop reports that verify completeness of collection/ processing activities

○ Deploy technologies that support collection activities. Capabilities to consider include:

  ○ In-place data processing

  ○ Data source integrations for conducting remote collections over the corporate network

  ○ “Spot” collector tools for collecting data from devices that aren’t connected to your network