Consilio / ADVANCED LEARNING INSTITUTE

Consilio Institute: White Paper

WHEN THE GAME IS AFOOT: INVESTIGATIONS AND EDISCOVERY

By Matthew Verga, Esq.
Director of Education

Consilio Institute: White Paper

WHEN THE GAME IS AFOOT: INVESTIGATIONS AND EDISCOVERY

TABLE OF CONTENTS

WHEN THE GAME IS AFOOT

The majority of eDiscovery work takes place in the context of litigation, but a significant amount of it takes place instead in the context of investigations. Although the available ESI and the available eDiscovery technologies are the same, the realities of handling investigations are different in some ways worth discussing.

In this paper, we will review what practitioners need to know about eDiscovery in the context of investigations, beginning with a discussion of common types of investigations and ESI's role in them. We will then discuss the need for speed and secrecy, the need for nuanced analysis and review, and the need to be prepared for later litigation.

Common Types of Investigations

Broadly speaking, there are three main categories of investigations in which corporations commonly become involved outside of, or before, actual litigation: (1) internal conduct or compliance investigations; (2) due diligence investigations; and, (3) regulatory enforcement investigations.

Internal Conduct or Compliance Investigations

Any organization of sufficient size will find it periodically needs to conduct an internal investigation of employee conduct or organizational compliance. Larger organizations and organizations in highly-regulated industries may find it to be a frequent occurrence. For example:

- Investigating alleged misconduct by an employee or manager (e.g., fraud or theft, data exfiltration, interpersonal misconduct)
- Investigating the circumstances surrounding the termination of an employee (e.g., was a termination for proper reasons properly documented)
- Investigating potential legal or regulatory violations in the conduct of business (e.g., was a particular transaction an attempt to manipulate a market)

> "'Come, Watson, come!' he cried. 'The game is afoot.'"
>
> - Arthur Conan Doyle, The Adventure of the Abbey Grange

Due Diligence Investigations

In the investigations context, due diligence refers most often to the evaluation of an organization for potential merger, acquisition, or partnership. Although financial aspects of such evaluations are generally the primary focus, evaluations of policies and procedures, of regulatory compliance, and of potential legal issues are of equal importance. Due

diligence in the mergers and acquisitions context is most common, but due diligence evaluations of prospective international partners have grown in frequency and importance as Foreign Corrupt Practices Act (FCPA) enforcement has accelerated in recent years.[1]

Regulatory Enforcement Investigations

The third category is regulatory enforcement investigations initiated by governmental agencies. These may result in no action, in settlements, in a regulatory hearing process, or in court, but before that, they are typically non-public investigations with discovery similar to that in litigation (though, one-sided in nature). The DOJ, the SEC, the FTC, the CFTC, the FERC, and more federal agencies all engage in such investigations, as do the equivalent state agencies and agencies of foreign governments.

This category includes the FCPA enforcement actions referenced above, as well as fraud investigations, trading investigations, and countless more subtypes. Another area of increasing enforcement activity is anti-money laundering (AML),[2] with two new statutes recently taking effect in the US (Anti-Money Laundering Act 2020 and the Corporate Transparency Act) and FinCEN and foreign enforcement agencies ramping up their activity.

ESI in Investigations

In each of the above scenarios, your organization will be faced with discovery needs similar to those you would face in litigation:

- ESI, including emails, electronic documents, and other materials, will almost certainly be relevant to your inquiry and necessary to meet your information needs
- As with litigation, those ESI materials are likely to be voluminous in scale and diverse in type and source
- Those materials will need to be preserved, collected, processed, managed, analyzed, and reviewed in an efficient and effective way

Thus, investigations share most of the same eDiscovery challenges as litigation, but investigations also entail added challenges:

1. First, investigations typically function on even shorter timelines than litigation discovery, making speed essential. Whether racing to assess internal risk or facing an agency-imposed deadline, time is of the essence. As Sherlock Holmes famously said: *"The game is afoot. Not a word! Into your clothes and come!"*
2. Second, because investigations potentially involve uncovering individuals' misconduct, controlling the dissemination of information about the investigation is important to prevent bad actors from intentionally spoliating materials or coordinating their stories.
3. Third, investigations often require more nuanced analysis and review of ESI for the same reason: bad actors don't always communicate about their bad actions in obvious and open language.
4. Finally, many investigation scenarios carry a strong possibility of formal litigation later, which means process decisions must balance investigation needs against being prepared for that potential litigation.

[1] Stanford Law School and Sullivan & Cromwell LLP, *Foreign Corrupt Practices Act Clearinghouse*, http://fcpa.stanford.edu/statistics-analytics.html (2022). Although filed enforcement actions decreased in 2021, this decrease is believed to be temporary and circumstantial, and enforcement is expected to return to pre-pandemic levels. See e.g. *FCPA Enforcement Actions Fall to Lowest Level in a Decade*, CORPORATE CRIME REPORTER (Jan. 8, 2022), available at https://www.corporatecrimereporter.com/news/200/fcpa-enforcement-actions-fall-to-lowest-level-in-a-decade/.

[2] *Money Laundering Enforcement Trends: Summer 2021*, MILLER & CHEVALIER (Aug. 9, 2021), available at https://www.millerchevalier.com/publication/money-laundering-enforcement-trends-summer-2021.

THE NEED FOR SPEED AND SECRECY

As noted above, investigations often function on even shorter timelines than litigation discovery and often require more-careful control of the flow of information.

Time pressure is great in most investigative scenarios. If you are working to assess an internal issue, you will want to identify and quantify risks to the organization as quickly as possible so that they can be appropriately mitigated. If you are working to respond to a regulatory agency's information request, you will likely be facing a tight, agency-imposed deadline, in a context where you want to satisfy the agency, in which the potential for renegotiation is limited, and in which the option to appeal to an independent judge is generally unavailable.

The need to more-carefully control the flow of information arises from the reality that in many investigative contexts you will be looking for bad actors within your own organization. Failure to control the flow of information (or to move with sufficient speed) provides opportunity for bad actors to spoliate evidence to cover their actions and to coordinate their stories with each other before talking to you. In either case, your ability to assess organization risks or to respond accurately to an investigating agency would be compromised, and the cost and effort required would become greater as you worked to overcome the attempted spoliation or penetrate the deliberate deception.

Achieving Speed

Speed in eDiscovery is a challenge in any context, but there are steps that can be taken at each phase to ensure things are moving as quickly as possible when you are racing an investigation deadline:

Before an Investigation Arises

- Improving organizational litigation readiness, including data mapping and data remediation, is the most effective way to speed up subsequent eDiscovery efforts of all types, including investigations
- The speed, efficiency, and reliability advantages that accrue from having less data, knowing where it all is, and having established procedures for how to act on it cannot be overstated

During Identification, Preservation, and Collection

- Undertake the brainstorming of potential sources immediately, then – rather than engaging in the more gradual processes that typically follow – move immediately to collect what you believe to be the key sources from the key custodians, so that preliminary analysis and review can begin while additional identification, preservation, and collection efforts are still ongoing in parallel
- Because, in this context, speed is most important, err on the side of over-collection from those key sources to avoid having to go back and conduct supplemental collections from those sources later

> **"I feel the need – the need for speed!"**
>
> - Jim Cash & Jack Epps, Jr., Top Gun

During Processing and Analysis

- When focused on maximizing speed, processing should be done without overly-complex or iterative filtering at that phase; standard de-NISTing, deduplication, etc., should all be done, but keyword and date filters should be saved for post-processing analysis in the review platform being used
- When processing data for time-sensitive investigations, it is beneficial to perform as a default step the necessary indexing to enable advanced analytic features (e.g., email threading, near-duplicate identification, conceptual search and clustering features, etc.) and technology-assisted review
- Advanced analytic features significantly increase the efficiency and effectiveness of analysis and assessment activities – particularly in early phases of an investigation when you may not be certain exactly what you're seeking

During Review and Production

- Unlike in litigation, where some concerns and complications persist, technology-assisted review can be used freely during internal investigations to significantly speed up needed review processes, and many federal agencies are now comfortable with investigation subjects using it as well (although methodology details generally have to be provided to the agency to secure approval)
- Finally, it is very common in the investigation context to engage in rolling productions over time, beginning with the key sources from the key custodians and moving through progressively lower priority materials as they can be completed; this can be a way to show a good-faith effort to cooperate when it is not possible to complete all needed work by an agency-imposed deadline

Controlling the Flow of Information

Controlling the flow of information actually goes hand-in-hand with achieving speed, as speed during identification, preservation, and collection is one of the best ways to stay ahead of the flow of information within your organization. As soon as a detailed hold notice (in the case of an agency preservation or production request) is issued, or as soon as active collection begins, any bad actors within your organization will be put on notice that you're looking for them and may take steps to destroy evidence or prepare for questioning. So, in situations where bad actors are suspected, certain collection steps may need to be taken before the hold notice is issued to all subject employees.

There are several collection strategies that can be employed to acquire key data without alerting suspected employees. For example:

- IT can typically collect from employees' active corporate accounts for email, messaging, documents, etc. without alerting the employees
- IT can take steps to preserve existing backups of sources as needed
- For an individual, a laptop or smartphone upgrade can be triggered, allowing IT to collect the current device and image it
- For an on-site team or department, IT can require all laptops be left at desks overnight for required security or software updates, and images can be made during those hours

In the context of an agency-directed investigation where a hold notice should be issued, the hold notice should still be created and issued immediately to relevant IT personnel, to other relevant systems owners within the organization, and to any relevant managers above the level of the suspected bad actor(s), with strict instructions about the confidentiality of the matter. Any unannounced collections, like the examples above, should happen as soon as possible thereafter, and then the hold should immediately be issued to the rest of the subject employees.

THE NEED FOR NUANCED ANALYSIS AND REVIEW

Analysis and review is the process of figuring out what happened by investigating your collected evidence, and that process is made more challenging when relevant individuals have actively tried to conceal what's happened – or at least tried to be subtle about it while it was happening. As we discussed above, investigations often concern misconduct of one kind or another, and most individuals instinctively try to render their own misconduct non-obvious.

It is not uncommon for individuals to communicate using euphemisms or coded language or to communicate using alternative channels. In some cases, individuals will attempt to obfuscate through the use of misleading file names or changes to file extensions. And, as we also noted above, some individuals will also attempt to destroy evidence if given the opportunity.

To overcome these challenges, the analysis and review process must be undertaken with these realities in mind, and it must be carried out in a way that will help you find these well-hidden needles in your haystacks.

Strategies for Nuanced Analysis and Review

Thankfully, there are a variety of tools and techniques you can employ to help you find hidden things and other unknown unknowns:

Random Sampling

One of the most useful tools for effective early case assessment, analysis, and review planning is random sampling. Random sampling is not susceptible to our biases and blind spots the way keyword searches are. It provides a cross-section of everything you have, including the unknown unknowns. It provides examples of the different kinds of language used by relevant individuals in different contexts, which in turn helps you identify language that stands out as atypical. If executed formally, it can even provide you with reliable estimates of how prevalent different types of material are in your overall collection.

Frequency Analysis

Many discovery tools make possible some form of frequency analysis. Rather than just testing various search strings and filters, frequency analysis tools show you all of the values of a particular type present in a given set of materials and tells you how frequently they occur. For example, you might be able to review a list of email correspondents for a particular custodian to learn who they correspond with most often and to look for any names that shouldn't be there. Or, a list of all frequently-occurring short phrases (e.g., 2-4 words) could be generated and reviewed to look for new potential search phrases and potentially-relevant euphemisms or coded language. As with random sampling, such analysis can reveal things for which you would not otherwise have known to look.

> **"It's like finding a needle in a stack of needles."**
>
> - Robert Rodat, Saving Private Ryan

Conceptual Analytics

Conceptual analytic tools include concept searching, concept clustering, and find-more-like-this features. Conceptual tools look at patterns of language rather than the specific language itself to facilitate searching and sorting that is more tied to contextual meaning than to precise phrasing. As with random sampling and frequency analysis, these tools are useful for finding things when you don't know exactly what you're seeking. Conceptual searching will return related results even if the exact words don't match; conceptual clustering can reveal topics you didn't know were there; and, find-more-like-this features can extrapolate from one relevant document – or even from a synthetic example of your creation – to find any similar documents in your collection.

Email Threading (and Multi-Source Coordination)

When investigating individuals' conduct, email threading features are very helpful for understanding the sequence of relevant communications and the individuals involved. Many discovery tools now offer visualization features built off of a threading analysis to let you map the flow of communication.

It is also often critical in investigations to be able to aggregate material from multiple communication sources in a chronological way. It is not uncommon for a conversation to begin in email, continue in an instant messaging or collaboration tool (e.g., Slack or Teams), and conclude in an OTC messaging application (e.g., WhatsApp). Only with all of the pieces in one place can the full sequence of events be reviewed.

Technology-Assisted Review

As we noted above, technology-assisted review can be used freely during internal investigations (and in many agency-initiated investigations). In addition to providing the speed advantage we discussed, technology-assisted review also extends the conceptual analytics advantage into your review process – taking your decisions and extrapolating from them based on semantic patterns rather than precise wording. Broadly speaking, TAR, comes in two varieties:

TAR 1.0 – Predictive Coding

- TAR 1.0 refers to the initial, categorization-based workflows offered in eDiscovery – many of which were, and are, referred to as predictive coding. Broadly speaking, these workflows involve leveraging a sampling process to create a training set or seed set (*i.e.*, a user-defined cluster or clusters), which the chosen software than uses to find other similar documents. These results are then reviewed and coded, and that coding is used to improve the software's results. This training cycle is iterated multiple times until an acceptable quality of results is achieved. The effectiveness of the whole process is measured using either a previously prepared control set or an additional random sampling effort.

TAR 2.0 – Continuous Active Learning

- TAR 2.0 refers to more recent workflows developed to leverage different mathematical approaches (*i.e.*, support vector machines and logistic regression). Rather than being based on identifying the similarities in a large, prepared training set like predictive coding, these workflows are characterized by continuous active learning that updates relevance scoring and prioritization for all documents dynamically as each additional document is coded by a reviewer.

  This is accomplished by focusing on a single, binary classification (*i.e.*, relevant to topic X and not relevant to topic X) and analyzing the differences in language between successive, single example documents to identify the hyperplane that best divides the relevant examples from the non-relevant examples on a multidimensional map. Each additional example the software analyzes and maps can lead the software to identify a more efficient

hyperplane between the two groups, improving its classifications.

These workflows emphasize speed over structure, and so, they work best in situations – like many investigations – where there is a clear, binary classification decision to make and where family groups and other contextual factors are less important than overall speed.

Forensic Analysis and Investigation

Finally, in the event that your analysis or review reveals gaps in the collection (or that you have other reasons to believe spoliation may have taken place), forensic analysts can undertake a variety of investigative steps to attempt to recover deleted materials or to determine user activities on a given device (e.g., to document data theft or destruction).

THE NEED TO BE PREPARED FOR LATER LITIGATION

As we noted at the beginning, time pressure is great in most investigative scenarios. Particularly when you are working to assess an internal issue, the desire to find out the facts and mitigate the risks can make speed feel like the overriding priority – a priority overriding even normal discovery processes. When it's just an internal investigation, why not skip forensic collection? Why not focus on fact-finding over process documentation? Why worry about preservation?

The Mess before the Mess

Despite the obvious incentives to rely on informal, ad hoc methods to handle the ESI discovery required for an internal investigation, there is a good reason not to succumb to the temptation: most investigative scenarios carry within them the potential for later litigation. For example:

- Investigating complaints about interpersonal misconduct leads to an employee's termination, which leads to a suit alleging wrongful termination or a suit from those co-workers affected by the misconduct
- Investigating questionable transactions leads to the discovery of employee regulatory violations, which leads to self-reporting to a regulatory agency, which leads to their formal investigation

> **"Well it's a mess, ain't it Sheriff?"**
>
> **"If it ain't, it'll do till the mess gets here."**
>
> - Joel and Ethan Coen, No Country for Old Men

- Investigating accounting irregularities leads to the discovery of employee fraud, which leads to the issuance of revised financial statements, which leads to shareholder lawsuits

The duty to preserve documents can arise before a case is actually filed or commenced, because the duty arises not when there is litigation but when there is reasonable anticipation of litigation (or agen-

cy action, etc.). As explained in "Guideline 1" of The Sedona Conference Commentary on Legal Holds, Second Edition: The Trigger & The Process[3]:

> A reasonable anticipation of litigation arises **when an organization is on notice of a credible probability that it will become involved in litigation**, seriously contemplates initiating litigation, or when it takes specific actions to commence litigation. [emphasis added]

Many of the same things that trigger an internal investigation could also count as notice of a credible probability of eventual litigation related to the underlying events. This will not be true in all investigative scenarios, but in the many where it is, a failure to look past the immediate information need to later legal obligations can result in inadvertent spoliation and other discovery complications.

Avoiding later sanctions and process challenges is a strong incentive to take a more formal approach from the beginning of many internal investigations.

Being Prepared for Later Litigation

To ensure preparedness for later litigation, there are three main things that need to be addressed during your internal investigation: preservation, forensic defensibility, and process documentation.

Preservation

The most important aspect of being prepared for potential later litigation is ensuring that relevant materials are not lost or destroyed. Even if the possibility of litigation is still too tenuous to have definitely triggered the duty to preserve, behaving as though it has will ensure relevant materials survive until you are sure. Taking steps to create backup copies of relevant email inboxes, file stores, or devices can be done even before you are sure you need to issue a hold or take other formal discovery steps. If the investigation reveals nothing, the backups can always be deleted later, but if the investigation reveals a serious issue, lost materials may not be recoverable later.

Forensic Defensibility

The next most important aspect of being prepared for potential later litigation is ensuring that the ESI materials you do collect in your investigation are collected in a forensically defensible manner. The desire for speed may make it tempting to review original files in situ or to have individuals forward relevant emails to you, but such actions can alter metadata, alter files, and create evidentiary problems in later litigation. Ensuring that your ESI is collected in a forensically-sound manner that preserves unaltered metadata, working with copies rather than original files, and documenting the chain of custody from the point of collection forward will ensure that your ESI evidence is ready for use in any later litigation.

Process Documentation

Finally, documenting your investigative process, decisions, and reasoning – particularly with regard to your preservation and collection steps – is invaluable in the event of later litigation. Sources are numerous and varied, ESI materials voluminous and complex, and investigations nonlinear. In an investigation of any size or length, you will not be able to recall all the details of what was collected and when, of what was preserved and why, or of what rationales supported other key decisions. Documenting these things as you go – as though in discovery – will ensure you have the records you need to effectively transition to formal discovery later or to explain your processes if challenged. And many later legal questions, such as those related to proportionality or reasonable efforts, can turn on the knowledge you had at the time.

[3] The Sedona Conference, *Commentary on Legal Holds, Second Edition: The Trigger & The Process*, 20 SEDONA CONF. J. 341 (2019), available at https://thesedonaconference.org/publication/Commentary_on_Legal_Holds.

KEY TAKEAWAYS

There are five key takeaways from this white paper to remember:

1. Investigations carry all of the eDiscovery challenges typical in litigation, with the potential added challenges of greater time pressure and intentional obfuscation or spoliation
2. Because speed and, at times, secrecy are critical in investigations, moving swiftly to undertake collection is critical and may need to be begun prior to hold distribution
3. Because most individuals instinctively try to render their own misconduct non-obvious, sampling, frequency analysis, and advanced analytic tools should be leveraged to help find relevant language and reveal unknown unknowns in collected documents
4. Smartphones, OTC messaging applications, collaboration tools, and other newer sources, have grown dramatically in popularity and importance for business communications, and they should not be overlooked when considering sources for an investigation
5. Despite the incentives to cut procedural corners during an investigation, the need to be prepared for potential litigation later means that preservation, forensic soundness, and process documentation should not be overlooked as you proceed

ABOUT THE AUTHOR

Matthew Verga is an attorney, consultant, and eDiscovery expert proficient at leveraging his legal experience, his technical knowledge, and his communication skills to make complex eDiscovery topics accessible to diverse audiences. A fifteen-year industry veteran, Matthew has worked across every phase of the EDRM and at every level, from the project trenches to enterprise program design. As Director of Education for Consilio, he leverages this background to produce engaging educational content to empower practitioners at all levels with knowledge they can use to improve their projects, their careers, and their organizations.

Matthew Verga, Esq.
Director of Education
m +1.704.582.2192
e matthew.verga@consilio.com
consilio.com