

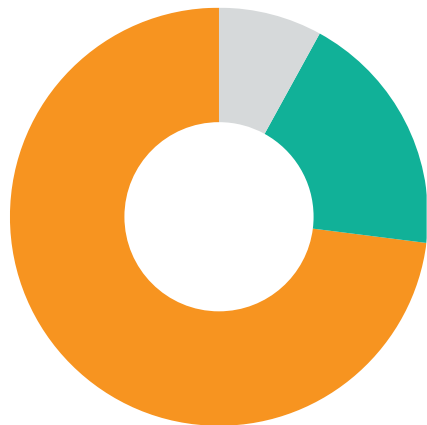
How to Achieve Discovery Workflow Nirvana

A New Approach to Efficient, Cost-Effective,
and Repeatable eDiscovery



The Costs of Inefficient eDiscovery Workflows

Review is the most expensive piece of the eDiscovery process.



eDiscovery Revenue Breakdown (\$7.5 Billion Market Size)

- **Collection**
8% (\$600M)
- **Processing**
19% (\$1.425B)
- **Review**
73% (\$5.475B)

The ultimate goal in any eDiscovery project is to make the right decisions about data at the right time. This enables corporations and law firms to save time and money throughout the eDiscovery process, especially during the review phase, the most expensive part of the process.

When it comes to document review, locating the relevant documents while culling out the non-relevant documents early on—without sacrificing quality—are the keys to saving time and money. However, to many corporations and law firms, achieving both simultaneously sounds impossible. Even with the increasing utilization of review platform analytics such as predictive coding, email threading, near-duplication analysis, visualization, clustering and other tools to help winnow down document collections (and thus drive down costs and increase efficiency), legal teams struggle to balance these goals. The result is a very expensive and inefficient review process.

The good news is that a more accurate, timely, and repeatable eDiscovery workflow is available right now that achieves both goals. Conduent Legal and Compliance Solutions’ technology-enabled, highly adaptive workflow replaces an inefficient and non-repeatable eDiscovery process with a quality-driven, connected and repeatable workflow that reduces costs and improves efficiency at every point in the review process, providing fewer false positives and improving quality control and accuracy. To understand how this adaptive workflow works—and how important it is to eDiscovery—let’s examine the high cost of review and how our innovative workflow drives those costs down, improves quality control and results in better and more predictable outcomes that improve with each new matter.

What’s Fundamentally Wrong?

It’s no secret that document review is the single most costly stage of eDiscovery. Take these staggering statistics, for example: In 2016, global eDiscovery market revenue was \$7.5 billion. Of that, collection accounted for 8% and processing for 19%. Review took up a whopping 73% share—\$5.5 billion. Furthermore, during the review phase, fewer than 10% of all documents reviewed for privilege in a single case are actually privilege—and 50% of privileged documents go unidentified. Finally, based on our experience, the average Fortune 1000 company reviews the same document at least 7 times and up to hundreds or more times across multiple matters.

As data volumes and sources grow exponentially, legal teams are struggling with the traditional way of doing things. Think of it as good news and bad news: the good news is that traditional analytics allow legal teams to save dramatic amount of money on review. The bad news is that true savings cannot be achieved with inefficient and non-repeatable workflows.

There are five major causes of inefficient workflows:

1. **Case-centric eDiscovery approach.** The typical approach to eDiscovery is matter by matter. That is, it is reactive and reinvents the eDiscovery wheel with each new case. Teams go through the exact same steps with each new matter: collect, process, analyze, review—yet it’s not transferrable across current or future matters.
2. **Limited ability to repurpose data across cases.** A case-centric approach limits the re-use of work product from previous matters to facilitate culling and/or coding decisions in new matters. It does not matter if analytics and reviewers make the same data decisions across fifty similar matters: it is technically impossible to automate the process using traditional workflow tools.
3. **Multiple stakeholders touch the same data.** In a typical case, data moves back and forth between multiple law firms, vendors, and eDiscovery platforms. Often, more than ten law firms are involved in a single matter. Knowledge from one case to the



Our unique adaptive workflow turns inefficiency on its head by reducing the cost and risk of review at every point in the process—not only for a single matter, but across all of a company's casework.

next, or even for a single case, is not easily (if at all) transferred between stakeholders, creating the risk of inconsistent coding and inadvertently exposing privilege or other sensitive documents.

4. **Inadequate quality control (QC).** Random sampling is the most common QC approach to evaluating review accuracy. However, random sampling alone does not efficiently focus QC efforts. By definition, a “random” sample is not necessarily a “representative” sample of the population at large, which makes it nearly impossible to draw meaningful conclusions from the results. Additionally, random sampling considers all documents equally, meaning a sample population is likely to contain a high volume of low risk documents that take unnecessary time, effort and money to review.
5. **Reactive vs. proactive.** Finally, typical eDiscovery workflows are reactive by nature. Basic approaches like deduplication and text searches are useful for first-pass culling, but only affect the data present in one case. Traditional workflows cannot proactively apply analysis from current and past cases to future matters to predict potential legal risk.

If you could implement a workflow designed to improve efficiency on current and future matters through the defensible reduction of objectively non-responsive content, enhanced privilege detection and the re-use of prior work product, it is possible to reduce both cost and risk—and this impact can be amplified with each new matter through ongoing heuristic analysis and machine learning that continually improve predictive results on future casework.

Beyond the Case-Based Approach: How to Achieve a Repeatable and Efficient Workflow

In a traditional case-centric workflow, legal teams would typically run—at the most basic level—tools such as de-NISTing, deduplication, and date and text searches. At a more advanced level, legal teams then might apply email threading, near-duplication, technology-assisted review (TAR) and other analytics. While basic and even more advanced filters reduce review populations, most resulting data sets typically retain a large ratio of non-responsive documents. Most often, the team proceeds with review anyway, because they have exhausted their standard filtering capabilities.

eDiscovery practitioners need not be doomed to repeat inefficient workflow processes. What if legal teams could accurately cull tens of thousands or more non-responsive documents from review populations—while also identifying and segmenting sensitive data, such as privileged documents—early on in the process? And what if they can take everything they learn about data in one case, and apply it across all subsequent cases? The result would be faster turnaround, lower costs and risks, better visibility and transparency at every point in the process, and more favorable outcomes.

Our unique adaptive workflow turns inefficiency on its head by reducing the cost and risk of review at every point in the process—not only for a single matter, but across all of a company's casework.

The adaptive workflow incorporates a variety of analytics-based features that can be delivered in a consistent and defensible manner. The problem is attacked from two sides: First, the approach focuses on enabling the defensible reduction of document review populations through the proactive identification of objectively non-responsive content. Second, the application of a best-of-breed technology stack enables greater control over review organization and prioritization efforts. By reducing the number of documents being reviewed and reducing the amount of time it takes to review those that remain, the adaptive workflow reduces overall review costs.



Beyond the application of traditional analytics, we have developed and implemented enhanced analytics to provide even more options for organizing and prioritizing reviews.

Let's contrast the traditional case-based scenario with this adaptive workflow, which is based on three major characteristics.

1. **Defensible reduction of the review population faster and earlier.** This adaptive workflow combines traditional and enhanced culling techniques, basic and advanced analytics, transparency and reporting, and human expertise in optimal order. The result is an efficient process that quickly reduces number of documents that are ingested into the review platform.
2. **Increased review efficiency.** This process increases efficiency by promoting responsive documents to the front of the review queue, accurately identifying privileged documents, reporting on search term precisions, accurately identifying non-responsive content, and organizing batches conceptually.
3. **Repeatable process with transferrable learning.** This workflow optimizes eDiscovery across multiple matters by applying learnings from past cases to future ones to save time and cost, as well as reduce future potential risk.

Components of an Optimized Review Workflow

Reducing costs starts with efficient and defensible culling that is also highly flexible. It's a simple equation: culling produces fewer documents, which takes less time to review, which reduces costs. Our approach does not end there. Advanced culling techniques learn from prior coding decisions, which it then applies to subsequent matters for greater workflow efficiencies.

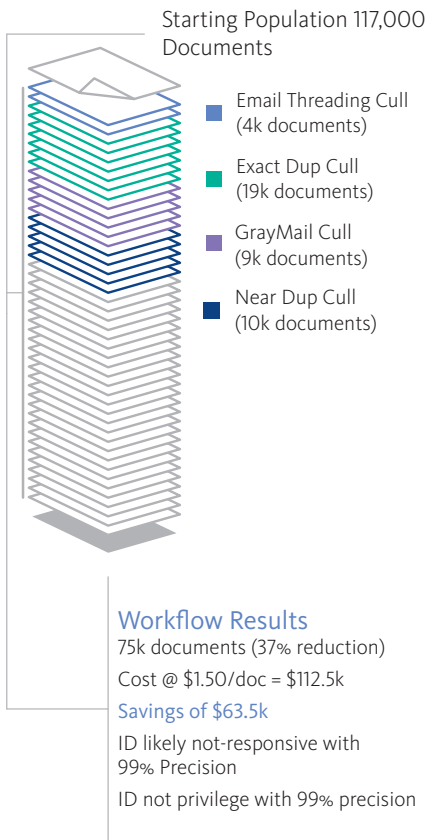
Flexible Culling

The adaptive workflow does not replace traditional filters to reduce large initial data sets; rather, it augments traditional tools for deeper and more accurate culling results. Basic approaches cull the initial document set with traditional filters such as date and text search terms, de-NISTing and deduplication. The adaptive workflow adds techniques aimed at eliminating redundant content and intelligently organizing the remaining review population for enhanced review efficiency. Email thread suppression, exact duplicate suppression and near duplicate suppression can further reduce a review population while ensuring that all potentially relevant content is still reviewed.

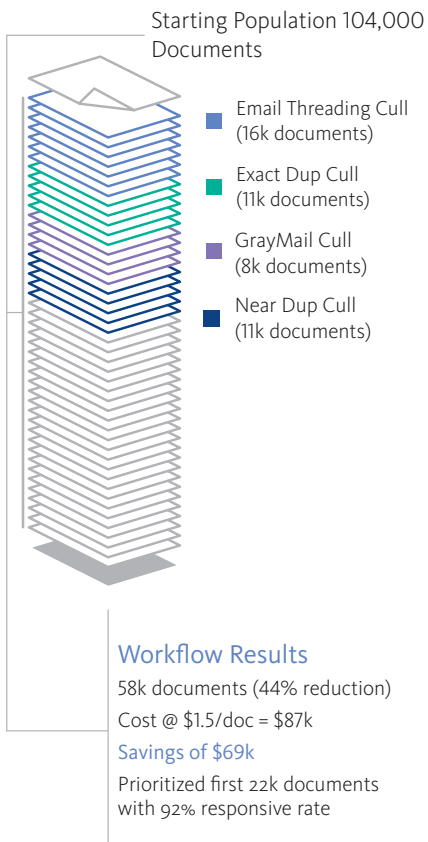
Email thread suppression leaves only messages containing unique textual content and attachments, while removing portions of email threads that contain redundant text. Exact duplicate identification moves beyond standard, hash-based deduplication methods and groups documents into exact duplicate sets based on 100 percent textual similarity. For example, if a PDF and a Microsoft Word document contain the same text, both would still survive a traditional deduplication approach; however, they would be grouped as exact duplicates enabling the potential to review only one version. Finally, near duplicate identification focuses on near duplicates with 99 percent textual similarity and greater (below that threshold meaningful discrepancies in textual content are noticed). The identification of this subset upfront allows legal teams to understand the impact of reviewing only pivot documents on overall review volume.

Beyond the application of traditional analytics, we have developed and implemented enhanced analytics to provide even more options for organizing and prioritizing reviews. The identification of Graymail isolates emails to which a custodian has "opted-in" or "subscribed." High frequency document detection (HFD) identifies content such as signature blocks, recurring memos, logos and weekly news letters – these subcategories are identified so review teams can decide how to handle each scenario. Finally, rules-based tagging is applied based on previous and ongoing heuristic analysis that associates specific document properties (such as number of recipients, time sent, subject, etc.) with review calls.

Case Study #1



Case Study #2



Objectively Non-Responsive Filtering

Most data associated with a legal matter contains a subset that can be classified as objectively non-responsive (ONR). These are documents that exhibit features of litigation content but will never be responsive, and which should be culled from review populations at the outset of the eDiscovery process. NIST filtering—commonly called deNISTing—is the most common approach to proactively remove documents without evidentiary value from potential review populations. Despite its limitations and overall impact on data reduction, in a traditional workflow, deNISTing typically is the only method used to identify and remove ONR content.

Graymail is email to which an individual has opted-in or subscribed to receive regular notifications (e.g., messages from retailers notifying an individual of deals, sales or specials). Unless a company has strong spam or Graymail filters, Graymail often represents up to 10 percent of most review populations—even after the application of date and term culling. A multi-pronged approach in this new workflow incorporates text and metadata to identify and cull Graymail from a population. Many forms of Graymail include specific text indicative of email that an individual has chosen to opt-in to receive (e.g., “click this link to unsubscribe”). Additionally, metadata characteristics including items like headers, senders, recipients and other properties are leveraged to distinguish Graymail from other potentially relevant email, such as emails that discuss business-related topics. Used together and in combination with other custom classification techniques, Graymail filtering quickly identifies and removes these categories of ONR data from a review population.

Identification of high frequency documents populations is another advanced ONR filtering technique that goes beyond traditional approaches. These are high-frequency documents such as signature blocks, recurring memos, logos and weekly newsletters. These items are typically not removed through deNISTing but should be culled from review sets as they are highly unlikely to contain relevant content.

Like the other filters, the application of rules-based tagging yields a more accurate and focused review set by using big data analytics combined with heuristic analysis to identify specific document properties that historically correlate to either relevant or not relevant review calls. For example, a certain sender and subject information known to be associated with privilege information, or a rule to flag documents with a specific sender domain if messages are frequently sent to more than 100 recipients, establishing a high confidence threshold that these messages are non-responsive.

Enhanced Privilege Classification

90 percent of all documents reviewed for privilege do not contain privileged information, resulting in a tedious, time-consuming and expensive process. The primary reason for poor precision is that traditional eDiscovery workflows use only text, which fails to take into account metadata, context of the conversation and other inputs, such as outside counsel lists.

Our approach dramatically improves the efficiency and lowers the cost of privilege review. Here’s how it works:

Through rules-based tagging, a combination of key document properties and privilege term analysis is applied to increase accuracy. For example, a rule to flag documents that contain certain sender and subject information known to be associated with privileged information can be applied. Then, by incorporating machine learning to analyze characteristics and key features of the data, the classification model separates privilege and non-privilege content. Specifically targeting privilege, text, metadata and lists of inside and outside counsel are used to train the models to search using context.



90 percent of all documents reviewed for privilege do not contain privileged information, resulting in a tedious, time-consuming and expensive process. Our approach dramatically improves the efficiency and lowers the cost of privilege review.

Context-based searching can detect if a conversation between two parties qualifies as privileged. For example, a conversation between two parties in which one of the parties has been identified as an attorney is identified as having a higher likelihood of containing privileged content than a conversation between two parties in which neither is an attorney. This contextual searching is impossible to do using traditional text-only search methods.

To better focus privilege review efforts, this new workflow moves beyond the simple binary classification of “potentially privileged” (where documents are either identified as potentially containing privileged information or not) by establishing a privilege classifier score. Similar to TAR-based relevancy scores, the privilege classifier scores range from 0.00 to 1.00 where a higher score means a document is more likely to contain privileged information. Additionally, documents are slotted into tiers based on previous and ongoing analysis of the scoring thresholds. For example, documents with scores greater than .30 are classified in a high probability tier; those with scores between .30 and .10 are classified as medium probability; and those receiving a score less than .10 are grouped into a low probability tier.

Based on client and test cases, this approach has routinely scored eight times better than standard keyword-based “privilege screen” approaches, resulting in a significant decrease in total review costs through reduced time investment in the privilege review phase. Legal teams can also use these techniques on a go-forward or look-back basis to gain greater insights into privilege identification efforts and associated costs.

Review Workflows Designed to Increase Efficiency

In addition to proactively identifying data sets that can be safely removed from review, thereby reducing review populations and associated costs, the adaptive workflow also increases efficiency. Our approach is designed to reduce the amount of time it takes to review what remains and to help focus efforts on documents more likely to be relevant.

Techniques aimed at increasing review efficiency include workflows for prioritizing responsive documents, accurately identifying privileged material, reporting on search term precision, the accurate identification of ONR data and workflows that facilitate the conceptual organization of review batches.

The prioritization of responsive documents is facilitated through access to advanced data visualization, including concept clusters and communications maps, the implementation of TAR 1.0 or 2.0 workflows and the conceptual organization of review batches. Advanced data visualizations allow review teams to quickly identify topics of interest (or junk) in unknown data sets. TAR 1.0 and 2.0 workflows take a small group of sample documents and identify likely responsive subsets within the remaining population. Also, by organizing review batches conceptually, review teams can achieve higher review rates by ensuring their reviewers are seeing similar conceptual content (and overall document format) document after document as opposed to more traditional batching approaches.

By accurately, and proactively, identifying likely privileged and ONR subsets, the adaptive workflow enables the intelligent organization of review. Documents with low privilege thresholds can be prioritized to facilitate quick and efficient review up front while those with high privilege thresholds can be sent through a separate, more focused review stream. Groups of documents identified as ONR can be de-prioritized, sampled against or sent through low-cost review streams. Furthermore, legal teams can review reports on search term precision so they can see what terms (responsive and/or privileged) are precisely returning expected results, or not. All of these related items and data points can be folded into QC workflows to further enhance efficiency and accuracy during downstream review as well.



By incorporating big data analytics, QC is substantially improved with each new case by identifying coding inconsistencies within groups of related items.

Application of Big Data Analytics

Now comes the heart of this new workflow, which gives users unique insight into their data across multiple matters, giving legal departments and law firms visibility into decisions across millions of documents or more.

This workflow applies big data analytics techniques to consolidate large eDiscovery document sets from multiple review platforms, repositories, vendors and/or law firms. The data analytics segments the data and automatically classifies and identifies data that needs to be reviewed—assessing up to billions of classification codes to identify documents that are relevant to new cases. This means that legal teams can eliminate re-reviewing the same documents across multiple cases—often up to 70 times or more—and substantially mitigate the risk of inadvertently exposing privileged or sensitive data.

This multi-matter approach achieves what a case-based workflow cannot. Legal teams can make connections across massive quantities of documents from prior cases across different platforms, automate document organization and assembly, and save significant time and money on future cases.

Benefits of Intelligent Workflow

Improved Quality Control

By incorporating big data analytics, QC is substantially improved with each new case by identifying coding inconsistencies within groups of related items. For example, quality control teams can quickly identify whether or not documents to be produced have items with high privilege classifier scores that warrant a second look. This, combined with reporting on search term precision, can enhance QC focus to prioritize documents with multiple inconsistency flags (such as a document with a high privilege classifier score that also contains a potentially privilege term with a high precision rating that was not coded as privileged and is slated to be produced).

Greater Transparency into All eDiscovery Matters

The last component of our workflow is a client-facing dashboard that enables legal teams to manage workflow across multiple platforms and matters, providing an unprecedented level of transparency across all cases. While many eDiscovery platforms have dashboards, this central management console consolidates data and project management across disparate platforms and matters for better decision-making.

The management portal aggregates case data from different platforms, analyzes metrics, and presents actionable reports to users. Typical reports include case status, project timelines, staffing, and expenses. Users can interact with real-time reports to review current case status and predict timeline and costs. They can also make active changes to reallocate resources, such as shifting review team assignments or signing in the same case to different platforms without losing visibility.

Conclusion

Our adaptive workflow offers a range of important benefits to law firms and corporations. In addition to increasing review speed by 35 percent or more based on client experience, law firms benefit themselves and their corporate clients with fast and flexible workflow propagate relevant data, enhanced QC, strategic data interactions, and the ability to prioritize or stratify review. Corporations benefit directly from enhanced business intelligence for more informed decisions to drive reduction of review costs, analytical models to enhance compliance efforts through the accumulation of work product, and reporting to assess the impact of cost reduction protocols.



Our worldwide footprint of data and review centers allows clients to quickly respond to discovery requests while remaining in compliance with data privacy and other legal requirements.

This new approach not only improves legal review; it also benefits compliance. The workflow and analytics can be configured to continually monitor incoming data for suspicious patterns that may indicate regulatory or compliance risk and accumulate work product, letting compliance and risk managers proactively identify compliance risk. This is a high-stakes capability in today's world of exploding data growth, increasing regulatory oversight, globalization, and an increasing reliance on partners and third-party vendors.

To benefit from this new workflow, law firms and corporations do not have to change their eDiscovery environment overnight. The new adaptive workflow protects existing investments, and users can customize the workflow to suit their automation comfort level. eDiscovery teams have every reason to adopt an approach that extends analytics and transparency across all matters, results in significantly lower costs and risks, grants better visibility and transparency in every point in the process, and leads to more favorable outcomes that improve with each matter.

Why Conduent

With close to \$1 billion invested annually in R&D, we employ more than 120 data scientists and have relationships with globally recognized research centers in the US and Europe. Through our team of legal subject matter experts and data scientists, we're constantly finding new ways to analyze legal data, making it more meaningful, and reducing costs while improving outcomes for our clients. Our worldwide footprint of data and review centers allows clients to quickly respond to discovery requests while remaining in compliance with data privacy and other legal requirements.

For more information on Conduent Legal and Compliance Solutions, visit us at: www.conduent.com/legalsolutions, or call 877.273.3887



Conduent Legal and Compliance Solutions ("Conduent") is not authorised to practice law, and neither offers legal advice nor provides legal services in any jurisdiction. The services offered by Conduent are limited to the non-legal, administrative aspects of document review and discovery projects. Conduent provides such services solely at the direction and under the supervision of its clients' authorised legal counsel.