White Paper

eDiscovery Best Practices in the Pharmaceutical Industry

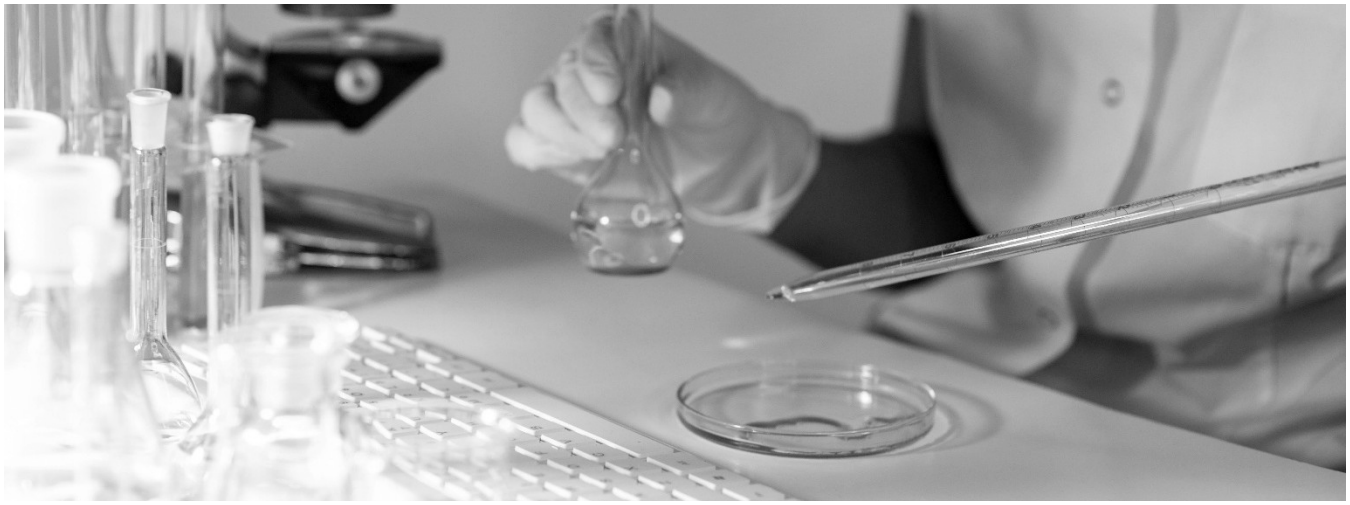
Phoebe Gebre • Claire Frazer • Miguel Villalobos • Robert Daniel • Scott Bien

I. Abstract

The pharmaceutical industry generates vast amounts of electronically stored information (ESI) across research and development, clinical trials, regulatory compliance, and corporate communications. Managing this data effectively during litigation, investigations, and regulatory inquiries presents unique challenges, including data volume, sensitivity, and jurisdictional constraints. This white paper explores the best practices for eDiscovery in the pharmaceutical sector, addressing legal and regulatory complexities, data governance strategies, and technological solutions. By implementing these best practices, pharmaceutical companies can enhance compliance, reduce costs, and mitigate risks associated with eDiscovery.

II. Problem Statement

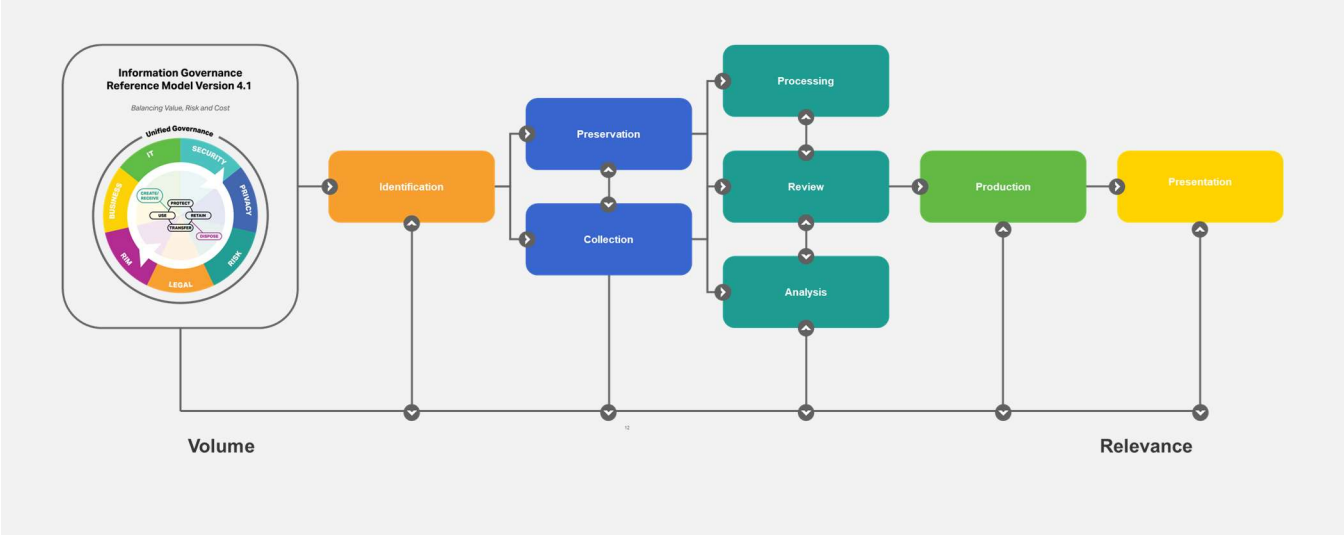
Pharmaceutical companies face a growing number of legal and regulatory challenges that necessitate a robust eDiscovery strategy. Regulatory agencies such as the Food and Drug Administration (FDA), European Medicines Agency (EMA), and Securities and Exchange Commission (SEC) impose strict data retention and production requirements, while privacy laws like Health Insurance Portability and Accountability Act (HIPAA) and General Data Protection Regulation (GDPR) add further complexity to handling sensitive patient and proprietary data. The industry also contends with massive, unstructured data repositories spanning clinical trial records, laboratory information management systems (LIMS), and corporate emails. Without a structured approach to eDiscovery, pharmaceutical firms risk non-compliance, legal penalties, and reputational damage.



III. Background

1. The Complexity of eDiscovery in Pharmaceuticals

The pharmaceutical industry is unique in its data landscape, as it must manage vast amounts of scientific, clinical, and commercial data across global operations. Electronic data sources include laboratory systems, regulatory submissions, email communications, and cloud-based collaboration tools. Data is often subject to strict regulatory oversight, with varying retention requirements depending on jurisdiction and data type. The high stakes of drug development and compliance investigations mean that failing to retrieve, review, and produce relevant data can lead to severe legal, financial, and reputational consequences.



2. Key eDiscovery Challenges

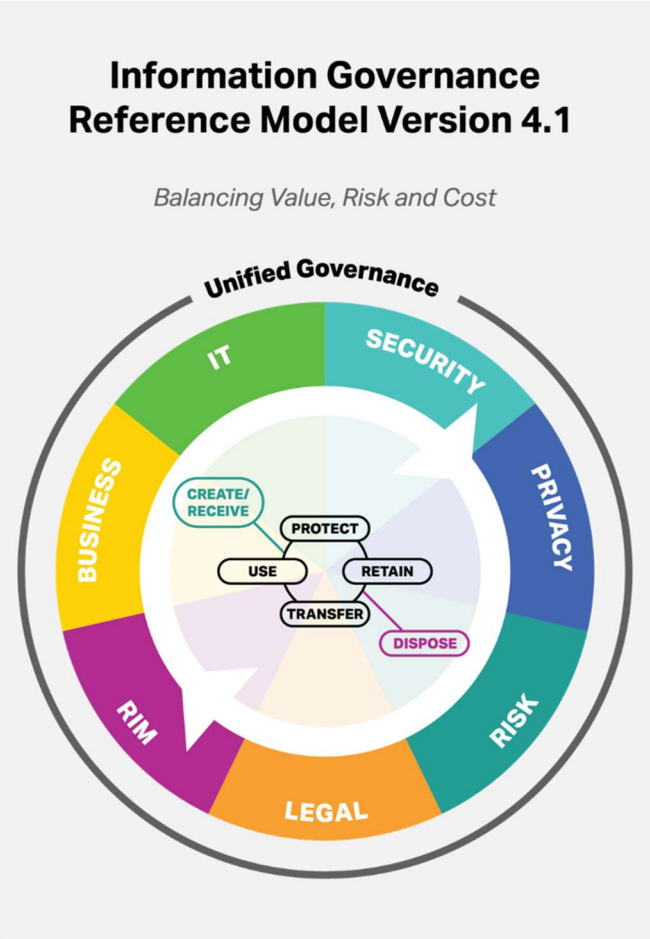
1. Data Volume and Fragmentation: Pharmaceutical companies must navigate large-scale data environments, often siloed across different departments and geographic locations.
2. Sensitive and Confidential Data: Data protection laws require strict safeguards for personal health data, trade secrets, and intellectual property.
3. Cross-Border Data Transfers: Companies must comply with international privacy laws when collecting and producing data for legal matters.
4. Regulatory Scrutiny: Agencies like the FDA and EMA require accurate, timely document production during audits and investigations.
5. Legacy Systems and Data Retention: Pharmaceutical firms often rely on legacy IT systems with inconsistent data retention policies, complicating eDiscovery efforts. Records may be stored in non-standard formats making them difficult to collect, process and review.
6. Complex Material and Technical Jargon: Document review requires resources well-versed in the subject matter and language of the content being reviewed.
7. Multi-Language Data Sources: Review and analysis can become more complex. Ensuring accurate, consistent review across languages requires advanced workflows.

IV. Solution: Best Practices for Pharmaceutical eDiscovery

1. Implementing Strong Information Governance

Pharmaceutical companies must establish comprehensive information governance (IG) policies to manage data effectively throughout its lifecycle. A proactive IG framework ensures compliance with regulatory requirements, enhances operational efficiency, and mitigates risks associated with litigation and regulatory inquiries Key Components of an Effective Information Governance Program:

- Data Classification and Organization: Establishing standardized taxonomies and metadata tagging to categorize data based on sensitivity, regulatory requirements, and business function. This enables faster retrieval and more efficient eDiscovery responses.
- Data Retention and Disposal Policies: Implementing industry-specific retention schedules aligned with regulations such as FDA 21 CFR Part 11, HIPAA, and GDPR. A defensible data disposal strategy reduces redundant, obsolete, or trivial (ROT) data, minimizing eDiscovery costs and security risks.
- Legal Hold and Preservation Processes: A robust legal hold program is a cornerstone for effective IG, ensuring that potentially relevant records are preserved in anticipation of litigation or regulatory reviews. By clearly defining the processes for issuing, tracking, implementing, and eventually releasing holds, organizations can manage data effectively and avoid costly risks associated with spoliation. Aligning retention and disposal activities with legal hold requirements enables organizations to distinguish between data that needs to be preserved and information that can safely be deleted.
- Departmental Coordination: Information governance must be a collaborative effort involving all stakeholders, including legal, compliance, IT, and business units. Establishing clear ownership of data governance policies ensures consistent enforcement across the organization.
- Audit Trails and Defensible Documentation: Maintaining detailed logs of data access, modifications, and preservation activities provides defensibility in court and regulatory reviews. Implementing governance, risk, and compliance (GRC) tools can automate this process and ensure regulatory alignment.



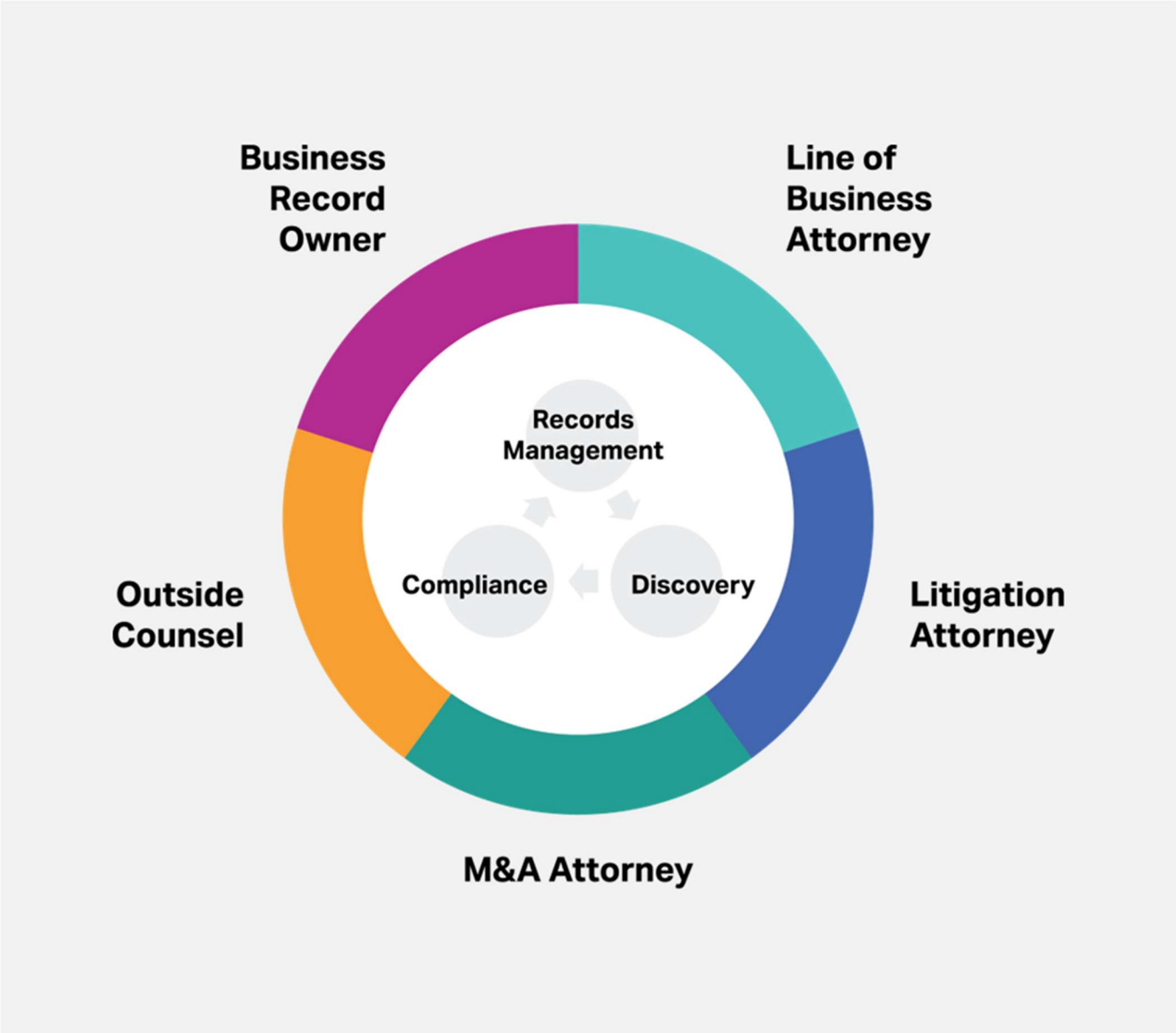
A well-structured data environment not only minimizes the burden of identifying and collecting relevant records during eDiscovery, but also enhances overall corporate data hygiene, reducing legal risks, operational inefficiencies, and potential sanctions for over-retaining personal information. Pharmaceutical companies that integrate proactive information governance with advanced data management technologies will be better positioned to handle legal and regulatory challenges with confidence.

2. Ensuring Compliance with Regulatory Requirements

A proactive eDiscovery strategy requires pharmaceutical companies to align their data management practices with stringent industry regulations governing electronic records, data privacy, and litigation procedures. Given the complexity of global compliance obligations, organizations must adopt a structured and defensible approach to managing electronically stored information (ESI) to reduce legal exposure and regulatory penalties.

Key Components of a Compliance-Driven eDiscovery Strategy:

- Regulatory Framework Alignment: Pharmaceutical companies must adhere to an intricate set of global and regional regulations, including:
    - FDA 21 CFR Part 11 (electronic records and signatures for regulated activities)
    - HIPAA (protection of personal health information in clinical trials and patient records)
    - GDPR and CCPA (cross-border data privacy and security compliance)
- EMA and MHRA Regulations (governing drug approval processes and compliance reporting in Europe and the U.K.)
- SEC and DOJ Requirements (applicable to financial disclosures, fraud investigations, and corporate compliance programs)
- Automated Compliance Monitoring and Audit Trails: Leveraging AI-driven compliance tools and automated audit logs ensures real-time tracking of data modifications, access patterns, and preservation actions. Audit trails serve as critical evidence to demonstrate regulatory adherence during government investigations or legal disputes.
- Defensible Chain of Custody: Implementing forensically sound data collection and preservation methods ensures that all ESI remains untampered and admissible in legal proceedings. Companies should document every step of data handling, from initial identification through collection, processing, and review.
- • Proactive Legal and IT Collaboration: Regulatory compliance cannot be siloed—cross-functional teams must work together to establish uniform compliance protocols. Legal, compliance, IT, and records management departments should conduct regular eDiscovery training, policy audits, and tabletop exercises to test response readiness for litigation holds and regulatory inquiries.
- Data Retention and Litigation Hold Enforcement: Companies must implement clear data retention and defensible deletion policies aligned with industry-specific regulatory mandates. When litigation or regulatory action is anticipated, litigation holds must be swiftly and effectively issued, acknowledged, and implemented across all relevant data sources, including emails, laboratory records, clinical trial databases, and structured enterprise systems.



A well-executed regulatory compliance strategy not only safeguards pharmaceutical companies against financial and legal penalties but also fosters greater operational efficiency, enabling faster, more accurate responses to regulatory and legal challenges. Organizations that integrate automated compliance tools, cross-functional collaboration, and defensible documentation into their eDiscovery workflows will be better equipped to navigate the ever-evolving regulatory landscape.

3. Managing Cross-Border eDiscovery Effectively

Pharmaceutical companies operating in multiple jurisdictions must navigate an intricate web of differing international data privacy obligations, data sovereignty laws, regulatory frameworks, and legal discovery obligations. This presents distinct challenges, and non-compliance can lead to substantial fines, legal penalties, and reputational damage. As a result, it is essential that organizations develop a strategic and legally defensible cross-border eDiscovery framework. It is good practice to align that framework with The Sedona Conference's International Principles on Discovery, Disclosure, and Data Protection, Second Edition (2017) (the International Litigation Principles), particularly Principle 1 (respect for foreign data protection laws) and Principle 2 (reasonableness and good faith) and The Sedona Conference Practical In-House Approaches for Cross-Border Discovery & Data Protection (2016) (the Practical In-house Approaches).

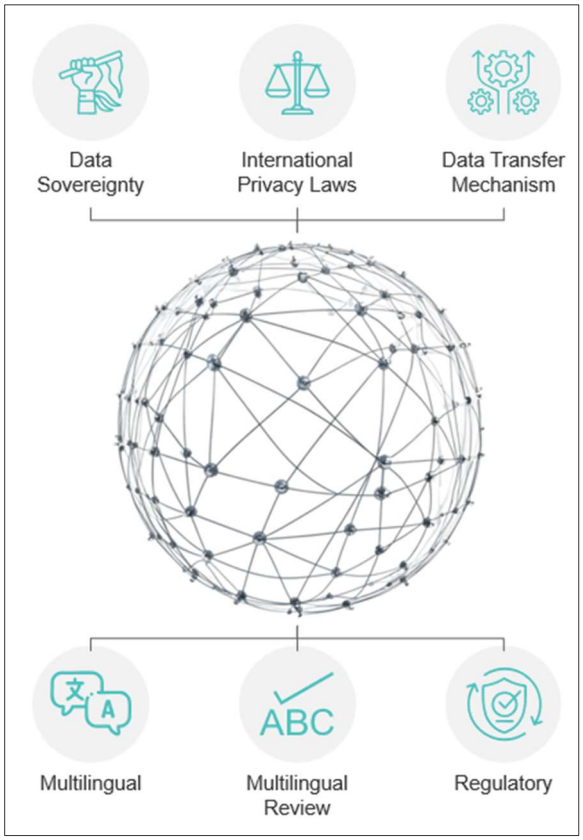
Key Challenges in Cross-Border eDiscovery

- Contrasting Definitions: In many jurisdictions, particularly in the EU, personal data is broadly defined and includes any information relating to an identifiable person, such as names, email addresses, or employment history (GDPR, Art. 4(1)). "Processing" includes collection, storage, review, and transfer, activities that may all be triggered during eDiscovery (GDPR, Art. 4(2)). In contrast, U.S. definitions are narrower, contributing to misunderstandings and compliance gaps during cross-border discovery.
- Conflicting Litigation & Privacy Obligations: Discovery obligations vary by jurisdiction. For example, U.S. litigation demands broad eDiscovery disclosures under the Federal Rules of Civil Procedure.

These obligations may conflict with the strict privacy obligations imposed in other jurisdictions such as the European Union (under GDPR), China (under the Personal Information Protection Law of the People's Republic of China (PIPL)), and Brazil's Lei Geral de Proteção de Dados Pessoais (LGPD). These obligations impose strict data privacy controls which dictate how personal and sensitive data, including employee and patient data, is collected, stored, and transferred, creating legal and operational tension.

- Regulatory Oversight and Investigations: Global regulators such as the U.S. DOJ, SEC, FDA, EMA, and U.K. Medicines and Healthcare Products Regulatory Agency (MHRA) frequently request cross-border data. As with the litigation discovery obligations mentioned above, companies must navigate the tensions between cooperating with regulators and complying with local data protection laws. The International Litigation Principles emphasize that companies should not only navigate these tensions but also document their efforts to do so under a standard of reasonableness and good faith. Also, early engagement with both U.S. regulators and local data protection authorities can demonstrate transparency and help manage expectations.
- Data Sovereignty and Localization Requirements: Many countries, including China, Russia, Brazil, and India, have enacted stringent data localization laws that mandate specific categories of data, such as health, biometric, or critical infrastructure-related information, be stored and processed domestically. These restrictions complicate centralized data collection and analysis for litigation or regulatory response. Where feasible, organizations should consider in-country or near-country review solutions and apply a tiered approach to production to avoid unnecessary transfer of data. Sedona Practice Point #5 stresses pre-collection planning, including identifying shared data systems and local data handlers.

Best Practices for Managing Cross-Border eDiscovery



- Data Mapping and Risk Assessment: Per Principle 6 of the International Litigation Principles, "Organizations should take good faith, reasonable efforts to retain, manage, and dispose of inactive data both on a prospective and retrospective basis." Pharmaceutical companies should conduct data mapping exercises to identify where sensitive data resides, who has access, which jurisdictional laws apply, and importantly when data should be disposed of to avoid the retention of data longer than is necessary. A risk-based approach ensures that legal teams proactively manage risk. This process should include IT, legal, compliance, and local business units to ensure a comprehensive understanding of both data locations and jurisdictional requirements.
- Localized Data Processing and Minimization: To comply with data localization laws, companies should implement in-region data hosting and processing solutions to reduce the need for cross-border transfers. This may include setting up regional data review centers or using secure, in-country cloud solutions that align with local compliance requirements. Organizations should consider staging document reviews, prioritizing U.S. custodians first to assess whether additional non-U.S. data is needed. This minimizes data transfer risks and supports proportionality principles. Establishing an in-country function can be logistically complex, particularly when navigating local employment and tax regulations, and so proactive operational planning as well as developing a well-defined operational framework in advance is essential.
- Adhering to Standard Contractual Clauses (SCCs) and Other Safeguards: Under GDPR, SCCs provide a legally recognized mechanism for transferring personal data outside the European Economic Area (EEA). Companies must ensure that SCCs are properly implemented and supplemented with additional safeguards, such as encryption and anonymization (Per Court of Justice of the European Union (CJEU) ruling, Data Protection Commissioner v Facebook Ireland Ltd and Maximilian Schrems (Schrems II) (2020)).
- Use of Data Anonymization and Pseudonymization: Before transferring data across jurisdictions, pharmaceutical companies should employ de-identification techniques to remove personally identifiable information (PII) while preserving the document's relevance for litigation or regulatory review.
- Regulatory Engagement and Country-Specific Protocols: Companies should establish clear, jurisdictional specific protocols for responding to legal and regulatory requests. This includes:

  - Templates and Process Documentation: In-house teams should maintain cross-border management templates (the International Litigation Principles, Appendix B: Cross-Border Discovery Management Template) to track obligations, legal justifications, data flows, and regulatory engagement actions for defensibility and auditability.
  - Engaging with local counsel to navigate country-specific laws and regulatory expectations.
  - Coordinating with data protection authorities where required before transferring regulated data.
  - Developing rapid response procedures for government inquiries, audits, and investigations enhances readiness and minimizes compliance delays.

- Legal and IT Collaboration for Secure Cross-Border Transfers: Organizations should implement secure data transfer mechanisms, such as encrypted transmission channels and zero-trust security models, to prevent unauthorized access or breaches during data transfers. Legal and IT teams should work together to establish protocols for defensible cross-border data handling. This integrated approach supports defensible data handling and aligns with The Sedona Conference's emphasis on coordination between legal and technical stakeholders, as emphasized in its commentary on practical in-house cross-border discovery approaches (the Practical In-House Approaches).

By implementing a proactive, structured and transparent approach to data governance and a risk-based compliance approach, grounded by the Sedona International Litigation Principles, pharmaceutical companies can balance their eDiscovery obligations with global data privacy requirements. This ensures defensible, efficient, and legally sound cross-border data management, supports cooperation with authorities, reduces the risk of regulatory violations and/or litigation exposure and ensures respect for individual privacy rights across jurisdictions.

4. Overview of the key Types of Litigation in the Pharmaceutical Industry

A. Product Liability Cases

Pharmaceutical product liability cases typically arise when a drug is alleged to cause harm to consumers due to defects, side effects, or inadequate warnings. These cases often include:

- Failure to Warn Claims: Allegations that the company did not provide sufficient warnings about potential side effects or risks.
- Design Defect Claims: Arguments that a drug's chemical composition or formulation is inherently dangerous.
- Manufacturing Defect Claims: Cases where a defect occurred during the drug production process, leading to contamination or improper dosage levels.

eDiscovery & ECA Considerations:

- Review of clinical trial data, pharmacovigilance reports, and adverse event reports submitted to regulatory agencies.
- Internal emails and meeting minutes discussing safety concerns or post-market surveillance data.

- Marketing and promotional materials that may misrepresent risks or benefits.
- FDA correspondence and regulatory approval records.

B. Patent Disputes (Hatch-Waxman Act Litigation & Biosimilars)

Patent disputes in the pharmaceutical industry are highly complex and often involve billions of dollars in revenue. These cases typically arise in two key areas:

- Hatch-Waxman Act Litigation (ANDA Disputes): When a generic drug manufacturer files an Abbreviated New Drug Application (ANDA) to produce a cheaper version of a branded drug, the brand-name company may file a patent infringement lawsuit to delay generic competition.
- Biosimilar Litigation: As biologic drugs become more common, companies file lawsuits under the Biologics Price Competition and Innovation Act (BPCIA) to prevent biosimilar competitors from entering the market.

eDiscovery & ECA Considerations:

- Patent filings, R&D documentation, and drug formulation records to establish infringement or non-infringement.
- Regulatory filings and correspondence with the FDA regarding exclusivity and market approval.
- Competitive intelligence reports and internal business strategy emails discussing the timing of generic entry.
- Third-party communications (e.g., with contract manufacturers or research partners) that may contain key admissions.

C. Regulatory Investigations (FDA, DOJ, FTC, and International Agencies)

Pharmaceutical companies are regularly scrutinized by regulatory agencies for compliance with drug safety, marketing practices, and pricing regulations. Investigations can be triggered by:

- Alleged misrepresentations in FDA filings or clinical trial misconduct.
- Off-label marketing claims, where a company promotes a drug for unapproved uses.
- Pricing and anti-competitive practices, including price-fixing and pay-for-delay agreements.
- Violations of Good Manufacturing Practices (GMP) leading to drug recalls or contamination concerns.

eDiscovery & ECA Considerations:

- Regulatory filings, internal compliance reports, and whistleblower communications.
- Emails and internal chat logs discussing marketing strategies or safety concerns.

- Board meeting minutes and executive communications regarding regulatory risks.
- Drug pricing strategies and communications with pharmacy benefit managers (PBMs).

D. Mass Torts & Class Actions

Mass torts and class action lawsuits are among the most significant legal threats to pharmaceutical companies, often resulting in billions of dollars in settlements and prolonged litigation. These cases arise when a large group of plaintiffs alleges harm from the same drug.

eDiscovery & ECA Considerations:

- Extensive document review due to the high volume of plaintiffs and claims.
- Use of predictive coding and AI tools to identify patterns in internal communications.
- Deposition preparation for key corporate witnesses, including scientists and executives.
- Public relations and reputational risk management, as these cases often attract media attention.

E. False Claims Act (FCA) and Whistleblower Lawsuits

Pharmaceutical companies often face False Claims Act (FCA) cases brought by whistleblowers (also known as qui tam lawsuits). These lawsuits allege fraud against government healthcare programs like Medicare and Medicaid. Common allegations include:

- Billing fraud – Overcharging government programs for drugs.
- Kickbacks to healthcare providers – Offering illegal incentives to doctors to prescribe certain drugs.
- Off-label marketing fraud – Promoting drugs for non-FDA-approved uses and seeking reimbursement from federal programs.

eDiscovery & ECA Considerations:

- Whistleblower emails, internal compliance audits, and financial records.
- Sales and marketing communications regarding incentives or off-label promotion.
- Medicare & Medicaid reimbursement documentation to assess potential fraud.
- Negotiations and settlements with government agencies that may impact legal strategy.

5. Streamlining Collection, Processing, Hosting, Review, Redaction, Production, and Privilege Logging

A well-structured eDiscovery workflow is critical for pharmaceutical companies to ensure efficiency, compliance, and defensibility in legal and regulatory matters. The complexity of pharmaceutical litigation and investigations demands a systematic, technology-driven approach to managing large volumes of electronically stored information (ESI). Each stage of the eDiscovery process—from collection through production—must be executed with precision to mitigate risk, reduce costs, and ensure regulatory compliance.

A. Collection

Targeted Data Collection

Effective data collection requires identifying, preserving, and extracting relevant information from key sources, including:

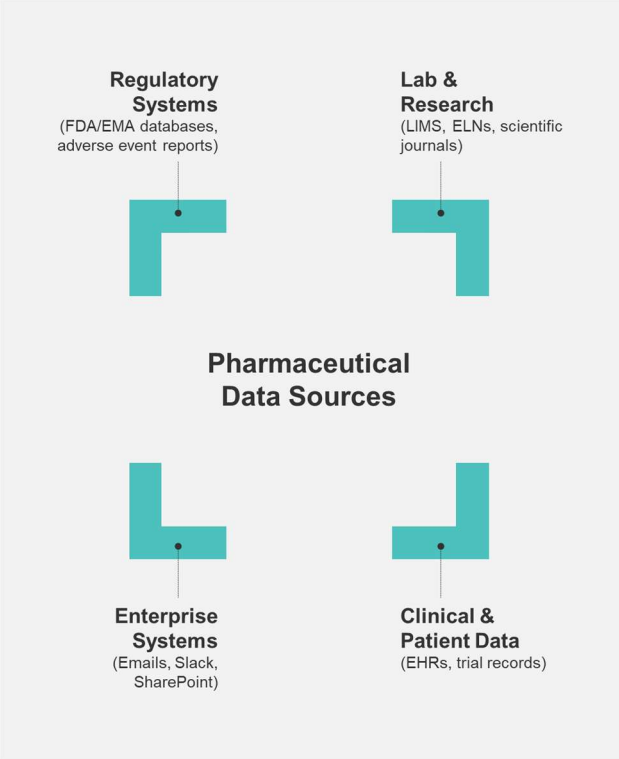
- Laboratory Information Management Systems (LIMS): Contains research, quality control, and testing data.
- Electronic Lab Notebooks (ELNs): Stores experiment records, formulations, and intellectual property.
- Regulatory Databases: FDA, EMA, and MHRA submissions, adverse event reports, and compliance records.
- Enterprise Communication Platforms: Emails, chat messages, and collaboration tools such as Microsoft Teams and Slack.
- Structured and Unstructured Data: Patient records, clinical trial data, and manufacturing logs.

A targeted collection strategy reduces the volume of unnecessary data and ensures relevant information is defensibly preserved without over-collection. Legal teams should work closely with IT and compliance teams to create data preservation notices and implement automated legal hold solutions to prevent data spoliation.

Forensic Data Collection Techniques

To ensure data integrity and admissibility in court or regulatory proceedings, pharmaceutical companies must utilize forensically sound data collection methods that:

- Preserve metadata (timestamps, file paths, authorship) to maintain chain of custody.
- Ensure defensibility by using industry-standard forensic tools (e.g., EnCase, FTK, Relativity Collect).
- Prevent data alteration by employing write-blocking techniques and hash value verification.
- Capture cloud-based and structured data sources while ensuring compliance with GDPR, HIPAA, and other regulations governing sensitive data.



A documented chain of custody must be maintained for all collected data to ensure its authenticity and reliability in legal proceedings.

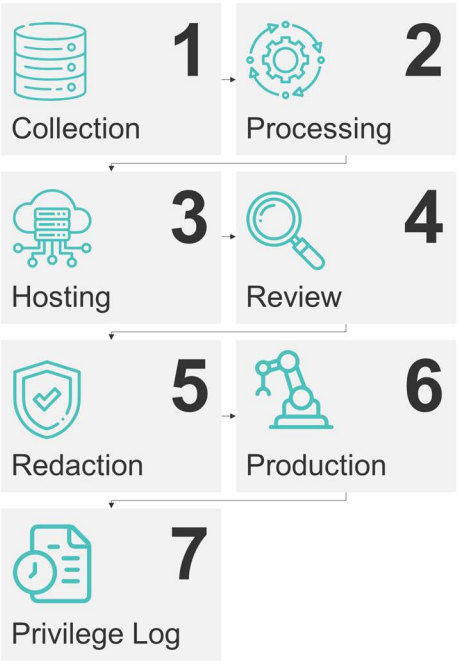
B. Data Processing and Hosting

Once collected, data must be processed to remove duplicates, filter out irrelevant content, and convert files into reviewable formats. This step involves:

- De-duplication and Near-Duplicate Analysis: Reducing redundant data to minimize review costs.
- Optical Character Recognition (OCR) and Indexing: Converting scanned documents and handwritten notes into searchable text.
- Metadata Extraction: Ensuring relevant metadata fields are preserved for legal analysis.
- Data Hosting in Secure Platforms: Utilizing secure, cloud-based eDiscovery platforms that comply with data security and privacy regulations.

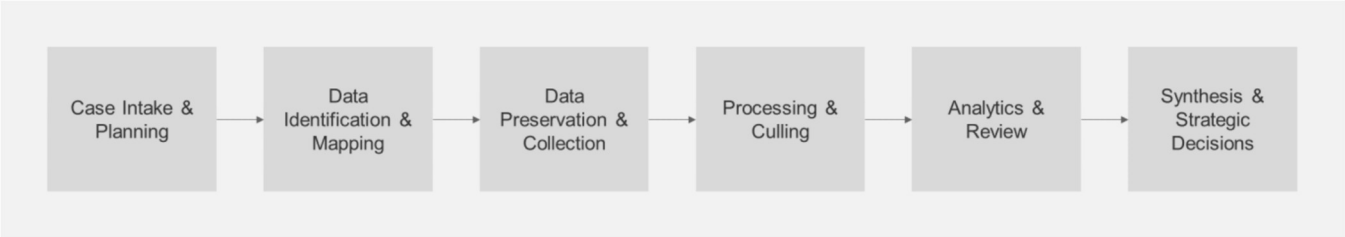
Pharmaceutical companies should leverage highly scalable hosting environments that allow for efficient searching, tagging, and retrieval of key documents.

C. Early Case Assessment (ECA)



i. Guiding Principles of the ECA Framework

- **Cross-Functional Collaboration:** Involve a multidisciplinary team from the start. This typically includes in-house counsel, outside counsel, eDiscovery specialists, IT/forensics, compliance, and business unit representatives. Early input from IT and data governance is crucial to identify data sources, avoid delays, and retention issues. Regular communication ensures legal strategy, technical processes, and business insights remain aligned.
- **Scalability and Repeatability**: The framework should handle matters of varying size and complexity—from a single-plaintiff dispute to multi-jurisdictional litigation. Core steps (data mapping, preservation, culling, analysis) remain the same, with the depth of analysis scaled to the case size. By documenting procedures and criteria at each phase, the process is repeatable and defensible, which is key for consistency across cases.
- **Global Consistency with Local Adaptation:** Maintain a unified approach to ECA globally while respecting jurisdictional differences. For example, discovery scope and privacy laws differ between the U.S. and EU. The framework calls for consulting local counsel and adapting to local rules (e.g., stricter data privacy or state secret laws) as part of ECA planning. Global coordination ensures that insights from one jurisdiction inform strategy in another, without violating local laws.
- **Technology-Agnostic, Analytics-Driven:** The methodology assumes use of standard eDiscovery technology (for processing, review, and analytics) but does not depend on any one platform. Leverage widely available analytics tools—concept-based clustering, advanced search, communication threading, timeline visualization—to expedite understanding of the data. These tools help cull noise and spotlight important documents early without manual review of every item. (For instance, concept clustering can group thematically similar documents, and email threading links conversations, regardless of the review platform used.)
- **Early Risk Assessment & Strategy Alignment:** ECA is fundamentally about risk management. From the outset, the team should define what questions need answering (e.g., "What's our potential exposure?", "Are there any 'smoking gun' emails?"). The framework emphasizes identifying key evidence and case themes within the first several weeks, allowing the legal team to develop a fact-based case strategy and budget forecast early on. All findings from ECA are continuously fed back to refine litigation strategy, settlement posture, and resource allocation.



1. Phase 1: Case Intake & Cross-Functional Planning

In this initial phase, the foundation is laid for a successful ECA by assembling the right team and information:

- **Case Profile & Team Assembly:** Define the basic case information—parties, forum(s), causes of action, and deadlines. Assemble the ECA team with representatives from legal (lead counsel, discovery attorneys), IT/forensics, compliance, and relevant business units. In a pharma context, this might include regulatory affairs or R&D personnel if the matter involves technical product issues. Establish points of contact and communication channels among these stakeholders.
- **Initial Strategy & Scope Definition:** Conduct a preliminary meeting to outline the alleged issues and potential scope of data. Align this with the broader litigation strategy set by trial counsel or the company (e.g., if the strategy is aggressive defense vs. early settlement, the ECA focus might differ). Define the objectives of ECA: for example, key facts to uncover, early case theory hypotheses, and what "go/no-go" decision points the organization is looking for. Clarify the expected timeline for the ECA (often the first 30–60 days of the matter) and any interim deliverables (like an initial case assessment memo).
- **Jurisdictional Considerations:** If litigation spans multiple jurisdictions (common in global pharma disputes), identify those at the start. Note where data custodians are located and any restrictions (e.g., GDPR in Europe, state secrets in China, blocking statutes in France). Develop a plan for cross-border data handling—such as obtaining data export permissions or deciding to host data in-region—upfront to avoid legal complications. Engaging local counsel early to navigate foreign discovery procedures is the best practice.
- **Communication & Preservation Plan:** Draft a communication plan to notify custodians and preserve relevant information. Issue legal hold notices to all key custodians identified, clearly describing the scope of data to retain (inclusive of emails, documents, lab notebooks, instant messages, etc., as appropriate to the case type). Ensure the hold is global in reach (i.e., sent in all jurisdictions involved and in multiple languages if needed). Set up a tracking system for custodians acknowledging holds and any issues reported (such as encrypted or legacy data needing special handling).

2. Phase 2: Data Mapping, Identification & Preservation (Global Data Considerations)

Global data mapping is essential in pharmaceutical cases, as data may reside in multiple countries and systems. In this phase, the team identifies where relevant Electronically Stored Information (ESI) resides and takes steps to preserve it, laying the groundwork for efficient collection:

- **Data Mapping:** Work closely with IT and records management to create a data map of all potential sources of relevant ESI. In pharma/life sciences, relevant data can be widespread—email servers, document management systems, clinical trial databases, lab information systems, shared drives, cloud collaboration tools (e.g., Microsoft Teams, Slack), and even physical archives for regulatory filings. Identify primary custodians (key employees) and data repositories for each custodian (laptops, network shares, cloud accounts, etc.). Crucially, record the geographical location of data stores, as this will affect how data can be collected and moved.
- **Preservation & Legal Holds:** Issue or update legal holds based on the data map findings. For each data source, coordinate with IT to ensure data is preserved (this may involve suspending auto-deletion policies, capturing snapshots of cloud data, etc.). For especially sensitive data types (clinical trial data, patient records, etc.), involve compliance/privacy officers to ensure preservation methods comply with regulations (e.g., HIPAA for patient health information, GDPR for EU personal data). All preservation actions should be logged for defensibility.
- **Interviews and Custodian Questionnaires:** Conduct custodian interviews to uncover additional information about data locations and case context. These interviews (or written questionnaires) often reveal informal data sources like personal devices, shared drives, or communication channels that aren't centrally known. They also help assess each person's role and potential knowledge of the issues. Document any leads (e.g., "Custodian X recalls an R&D Slack channel about the project") and incorporate those into the data mapping.
- **Global Privacy and Transfer Considerations:** When data is located in jurisdictions with strict data laws, plan accordingly. For example, EU personal data may require anonymization or special handling before review. China may restrict export of data without review by authorities. The ECA team should segment data by jurisdiction, keeping data local as required during ECA (e.g., using review servers in-region or reviewing on-site) to remain compliant. If data must be transferred, work on obtaining necessary approvals or use of data transfer mechanisms (standard contractual clauses, etc.). These considerations may slightly slow down data access, so build them into the ECA timeline.
- **Defensible Process Documentation:** Throughout identification and preservation, maintain clear documentation. This includes lists of data sources, custodians, hold notice details, and descriptions of any data excluded from preservation (with rationale). A defensible audit trail will show that the company took reasonable steps early to identify and safeguard relevant information, which is crucial if preservation is later challenged in court.

3. Phase 3: Early Data Collection, Processing & Culling

Once key data sources are identified and preserved, the next step is to collect and process a data subset for early analysis. The goal is to quickly reduce the data to a manageable volume for ECA review without eliminating potentially important information:

- **Targeted Collection:** Work with forensics/IT to collect data from the most relevant custodians and systems first. This might mean imaging email accounts and work laptops of a core team involved in the events. Collection at this stage can be iterative—for example, start with 5-10 key custodians most likely to have critical documents, then expand if needed. Ensure collections are done in a forensically sound manner (preserving metadata, using chain-of-custody forms). For cloud data and databases, use available export tools or APIs to pull data (e.g., export Slack conversations, database reports).
- **Processing & Early Filtering:** Ingest the collected ESI into the eDiscovery platform for processing (extracting text, metadata, de-duplication, etc.). It is often more cost effective to upload only text and metadata post processing for ECA purposes. Apply standard culling techniques to drop clearly irrelevant data early:
    - De-duplication and near-duplication detection to remove duplicate files and identify clusters of very similar documents.
    - Date range filters focusing on the period relevant to the dispute (e.g., start from a year before the earliest complained-of event through a cutoff).
    - File type and size filters if appropriate (e.g., excluding system files or audio files if not relevant, while being careful not to toss potentially relevant data like voicemails if they matter).
    - Domain or keyword filters to eliminate obvious non-responsive material. For example, filter out domains of newsletters or personal email (gmail.com) unless known to be case-related. Using a keyword like “unsubscribe” or common social words can help isolate personal or junk emails. (One approach: search for trivial content like lunch invitations, then cull entire email threads of that nature in bulk.)
- **Leverage Analytics for Culling:** Beyond basic filters, use analytics to aid early culling. For instance, some eDiscovery tools provide email threading and spam detection that can automatically cluster and tag mass-email threads (newsletters, etc.) for removal. Concept clustering might help identify a cluster of documents that are unrelated to the case issues (e.g., an unrelated project) which can then be set aside. These analytics-driven cuts reduce data volume while preserving a copy of everything in the platform, so nothing is lost if you need to retrieve it later.
- **Sampling and Validation:** After culling, perform quality checks. Sample a random set of excluded documents to verify that nothing obviously relevant was accidentally filtered out—this validates the culling criteria and demonstrates defensibility. Likewise, sample from the retained data to get a sense of relevance density (e.g., what percentage looks likely relevant) to inform the next steps and refine search terms.

- **Result:** By the end of this phase, you should have a significantly narrowed ECA data set that is indexed, deduplicated, and ready for focused analysis. The culling approach should be documented (search terms used, date filters applied, etc.) in case you need to justify why certain data was not reviewed as part of ECA.

4. Phase 4: Analytics-Driven Early Case Analysis

With a manageable data set in hand, the team now conducts a deep-dive analysis using advanced analytics and targeted review. The aim is to surface key documents, communications, and themes efficiently, rather than linearly reviewing everything. Best practices in discovery analytics are employed here:

- **Concept Clustering and Thematic Analysis:** Utilize concept clustering to have the review platform group documents by conceptual similarity. This unsupervised machine learning approach quickly reveals major themes in the data without preconceived keywords. For example, documents might cluster into groups about "clinical trial results," "patent filings," "manufacturing issues," etc., giving an immediate birds-eye view of what topics are present. Reviewers can prioritize clusters that are likely important to the case. Concept clustering has been shown to dramatically speed up insight-gathering (in one instance, finding crucial evidence in days instead of weeks), and can cut review time significantly.
- **Email Threading & Conversational Analytics:** Apply email threading to reconstruct entire email conversations. Threading links all replies and forwards, allowing the reviewer to read one coherent chain from first email to last. This provides context (who said what, and when) and prevents reviewing duplicate content multiple times. It also sheds light on communication patterns (for instance, seeing who is included or dropped in later replies). Benefits: The team can quickly understand the tone and content of discussions (e.g., identifying if a thread shows internal concern about a drug's side effects vs. a routine update). In addition to email, many platforms now support the threading of messages from chat applications (Slack, Microsoft Teams, etc.). Short messages are presented in a chat transcript view, preserving emojis or reactions, which helps interpret the meaning. By threading these, the ECA team can follow important discussions in context rather than fragmented messages.
- **Advanced Search and AI:** While clustering and threading let patterns emerge, traditional search is also used in tandem for specific queries. The team should run iterative keyword searches (including synonyms, product code names, etc.) to pinpoint documents on central issues. Conceptual search

can find documents "like" a known important document even if they don't share keywords. Tools with Generative AI based conversational inquiry can often lead to high level insights quickly by asking natural language questions to the system about key people, events and facts.

- **Many tools offer machine learning/predictive coding (TAR)** which, even if not used for full review, can rank documents by likely relevance during ECA. By training a quick model on a handful of relevant documents, the team can surface other high-scoring documents early (a form of continuous learning to home in on important evidence). This initial analysis can also be repurposed to save time and costs during the review phase.
- **Data Visualization and Communication Mapping:** Leverage visualization features to complement textual analysis. Communication mapping (a form of social network analysis) can show who communicated most with whom during the relevant period—useful in a big case to identify additional key players or unusual communication patterns (e.g., a normally uninvolved executive suddenly in email loops during a crisis). Timeline analysis tools plot document counts or key events over time, helping pinpoint spikes (for instance, a surge of emails around a drug recall date). These visual insights guide where to dig deeper.
- **Issue Tagging and Preliminary Review:** As important documents are found, attorneys should tag and annotate them by issue (e.g., "design defect", "causation", "compliance red flag", "damages") to build an evidence outline. They should also flag privileged communications to assess how much of the story might be hidden behind privilege (and to plan for privilege logs). The ECA team may perform a quick first-pass review of a sample of documents from each major cluster or custodian to summarize what that subset contains. The idea is not to review everything, but to extract key information. By the end of this phase, the team will have a set of hot documents, a sense of overall data themes, and initial answers to the big questions (e.g., "We have found emails suggesting early knowledge of the issue", or "Thus far, no evidence of the alleged patent infringement in the design documents"), All these findings are captured for the next phase.

5. Phase 5: Synthesis of Findings and Strategic Alignment

In the final phase of ECA, the team compiles the insights gained and uses them to inform case strategy and the next steps in the litigation:

- **Key Findings Report:** Summarize the critical documents and facts uncovered. This report or memo should outline the main case themes identified, organized by issue. For instance: "Product Quality Issue: Internal emails in March 2020 show discussion of an increase in adverse events (Document IDs…)", "Patent Validity: Lab notebooks from 2015 indicate the invention was reduced to practice later than claimed (Document IDs…)", etc. Including a brief description and significance of each hot document is helpful. This equips counsel with a quick reference of the evidence landscape discovered so far.
- **Preliminary Case Assessment:** Evaluate the strengths, weaknesses, and risks for the case in light of the ECA findings. Questions to address include: How strong is our evidence on key points? Are there damaging documents (the "bad facts") and what is their likely impact? What important information is

still missing or uncertain? This assessment connects the evidence to legal elements of claims/defenses. It should also identify any factual issues that could spawn early motion practice (e.g., grounds for a possible motion to dismiss or need for an injunction). Essentially, the team updates the initial case strategy with reality-checked insights.

- **Decision Point – Settle or Proceed:** If the ECA has revealed high risk (e.g., a "smoking gun" document proving liability), the organization may decide to pursue early settlement or other resolution. Conversely, strong evidence in your favor might embolden a vigorous defense or a countersuit. Because ECA's key function is to enable informed decisions on case direction, the findings should be presented to litigation leadership and business stakeholders to make that call. Often this involves a meeting or presentation where discovery attorneys walk through the evidence and scenarios (sometimes using a few key documents as examples).
- **Aligning Global Strategy:** In global matters, ensure that the ECA insights are shared with counsel in other jurisdictions (in a manner consistent with local privacy or secrecy laws). For example, an internal investigation finding in the U.S. might need to be carefully summarized (not raw data) before sharing with EU counsel due to privacy restrictions. The goal is a coordinated global defense or prosecution strategy—understanding how evidence found in one country might affect proceedings in another. Also, consider jurisdiction-specific needs: if a regulator in Country X is focused on a certain issue, highlight any findings on that to the team handling that investigation.
- **Next Steps & Workflow Transition:** Plan the transition from ECA to full discovery or case development. This includes identifying additional data or custodians should be collected based on what ECA showed (e.g., "We need to collect the marketing team's files, as ECA suggests their role is bigger than initially thought"). It also involves setting review workflows for the broader review: ECA can inform how to prioritize review batches (perhaps by focusing on the clusters that had the most important documents first). If using external review teams, the ECA results and tagging can be handed off to guide them. Additionally, outline any further analytical tasks—for instance, if during ECA you found a complex set of scientific reports, you might plan to bring in an expert to interpret them in depth.
- **Documentation and Defensibility:** Finally, compile the documentation of the ECA process (custodian lists, search terms, analytics used, etc.) into the case file. Not only is this useful for internal knowledge (especially if personnel change), but it also shows, if needed later in court, that the ECA was done methodically and in good faith. Many jurisdictions encourage early discovery meetings; the ECA results will prepare you to engage in meet-and-confer negotiations knowledgeably (e.g., you can better discuss which data sources are truly relevant or how costly certain discovery would be, potentially convincing the other side or the court to limit scope).

ii. Adapting the Framework to Different Case Types

The above ECA framework is meant to be flexible and adaptable. While the core steps remain consistent, specific litigation types in pharma/life sciences have unique considerations:

- Product Liability Cases: These often involve large volumes of data due to multiple plaintiffs or multidistrict litigation (MDL). Focus on aggregating and analyzing adverse event reports, safety databases, and complaint files. Use concept clustering to group similar incident documents or customer complaints, which can reveal patterns in alleged injuries or product issues. Involve pharmacovigilance experts early to interpret technical drug safety information. Also, there will likely be a need for historical regulatory submission data (to see what the company disclosed to regulators about risks). An ECA in a product liability matter should quickly identify any documents suggesting the company knew of a defect or risk and how it responded—these will drive the case's risk assessment. Workflows should account for coordinating ECA across many cases (if part of an MDL or global litigation campaign, consider a centralized ECA team that feeds findings into all cases for consistency).
- Intellectual Property & Patent Litigation: These matters hinge on technical details and dates. Emphasize collecting R&D documents, lab notebooks, patent prosecution files, and communications with patent offices. Expect a significant subset of documents to be highly technical—consider bringing in a technical expert or patent agent as part of the ECA team to help identify key inventions, prior art, or disclosure issues in the documents. Use analytics to correlate scientific terms with inventor communications (concept clustering might separate documents by subject such as "formulation stability" or "synthesis method"). Email threading is useful to trace how an invention was discussed internally over time, or how trade secrets might have been shared. Be mindful of privilege: patent attorneys' communications and opinion letters should be identified and segregated early. Also, in global pharma IP disputes, multiple jurisdictions (U.S., EU, Asia) might each have related patent cases—coordinate ECA to gather a global view of the innovation timeline and any public disclosures, as this will inform strategy in all venues.
- **Regulatory Investigations and Enforcement:** ECA in an investigation (e.g., by the FDA, DOJ, or a foreign regulator) often operates under tight deadlines and a need for utmost thoroughness. The case might not have a formal complaint but rather an investigative demand or subpoena. Prioritize data related to communications with the regulator, internal compliance reports, audit findings, and the specific subject matter (e.g., if it's a manufacturing compliance investigation, focus on quality assurance (QA) and quality control (QC) documents, batch records, and emails around deviations). Use analytics like keyword expansion to catch code words employees might use to refer to problematic topics. Since regulators may request explanations quickly, use the ECA to assemble a chronology of events and document sets that tell the story of what happened and who knew what when. Jurisdictional nuance: if multiple countries' regulators are investigating (say an FDA inquiry and an EU EMA inquiry), ensure findings are packaged separately according to what data can be shared, but internally maintain a composite understanding. The framework should remain platform-

agnostic but assume you'll use standard eDiscovery tools to rapidly search and produce data for the government. One must also prepare for the possibility of a report-out: summarizing findings to the regulator—the ECA analysis will form the backbone of that report.

- **Commercial & Contract Disputes:** These involve business agreements (supply contracts, partnerships, acquisitions, etc.) and often hinge on interpreting what was agreed and whether obligations were met. The relevant data sets are usually smaller (focused on contract documents and the communications of the negotiators or managers of the deal). Here, document clustering can help group communications by topic (e.g., "pricing discussions" vs. "delivery timeline emails"). Pay special attention to draft versions of contracts and related email threads, as differences may show representations or promises made. The ECA should aim to quickly identify any written evidence that supports each side's interpretation of the contract. If it's a licensing dispute in life sciences, ensure you gather communication with the other party over the license term (to see course of performance). Because these disputes may involve financial data as well, consider including finance team members to ECA if needed (to interpret revenue records or damages-related info). Global consideration: commercial disputes can cross borders (e.g., international supply agreements), so be mindful of where contract data is stored and any need to translate documents that are in other languages as part of ECA.
- **Consumer Health Product Litigation:** Consumer health cases (e.g., involving over-the-counter medicines, supplements, or medical devices sold directly to consumers) often resemble product liability but usually emphasize marketing practices and consumer communications. In ECA, focus on collecting marketing materials, labeling claims, advertising approvals, and any consumer feedback (complaints, call center logs, social media interactions). Use analytics to cluster documents related to marketing campaigns or product launch materials—this can show what claims were made and when. Also, search communications for discussions about how risks or product benefits were portrayed to consumers. The team should include regulatory or marketing compliance experts to identify any deviations from approved language (for instance, off-label promotion evidence). If the litigation is a consumer class action (say, false advertising), early case assessment should evaluate the consistency (or inconsistency) of messaging in public materials versus internal knowledge. Jurisdictionally, consumer protection laws vary (EU vs U.S.), so if the case spans regions, the ECA should note any documents that might be problematic under one country's consumer laws even if not under another's. The outcome of ECA will guide whether to fight the allegations or seek an early settlement, especially if a problematic email like "we know this claim is unsubstantiated" is found.

Implementing a structured ECA framework enables pharmaceutical and life sciences companies to manage complex litigations and investigations proactively. By rapidly distilling large, disparate data sets into key insights using advanced analytics, cross-functional expertise, and globally aware practices, discovery teams can drive case strategy rather than react to it. This high-level, platform-agnostic approach ensures that regardless of the litigation type or jurisdiction, the organization approaches the matter with an informed plan, controlled costs, and a cohesive strategy. ECA, when done thoroughly and consistently, not only reduces the data review burden but also provides a strategic advantage—

empowering attorneys to negotiate, litigate, and resolve cases with confidence backed by data-driven knowledge. By following this framework, legal teams in the life sciences sector can navigate the unique challenges of their cases while maintaining alignment with broader global litigation goals and compliance obligations.

D. Document Review

Document review is a critical component of legal discovery and compliance processes. As the most expensive part of the EDRM lifecycle, it requires both efficiency and accuracy to manage the massive volumes of data that need to be reviewed in modern litigation and investigations in the pharmaceutical industry.

The difference between a smooth, defensible review and production and a problematic one often hinges on the presence—or absence—of well-designed, repeatable workflows. These processes form the backbone of consistent, high-quality reviews by establishing clear, documented methodologies that apply regardless of which outside counsel is engaged or how the project management and review teams are composed.

When document review teams operate with standardized procedures, they create an audit trail that reflects thoroughness and a good faith effort to meet discovery obligations. This systematic approach becomes particularly valuable when the adequacy or completeness of a production is challenged.

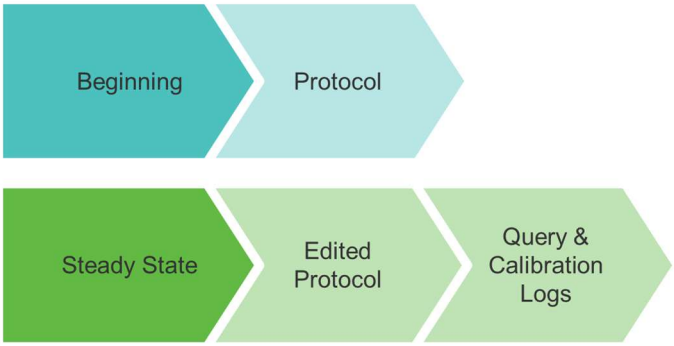
Conversely, the absence of repeatable processes introduces significant risk. Inconsistent review standards can lead to disparate treatment of similar documents, inadvertent disclosure of privileged materials, or the misclassification of relevant documents as non-responsive. These inconsistencies can compromise the integrity of the review, expose the corporation to sanctions or adverse inferences, and necessitate costly re-reviews.

i. Repeatable Processes and Project Management

An effective document review team functions as a force multiplier—striving to replicate, as closely as possible, the decisions that senior attorneys familiar with the case would make if they had the time and opportunity to review every document themselves. A best-practice review framework enables counsel's strategic guidance to be applied consistently across a team of reviewers by embedding it into a scalable, structured decision-making process. This structure helps ensure that individual reviewer decisions remain aligned, calibrated, and defensible.

Document review centers on two key objectives:

1. To identify all documents that are relevant to the matter and responsive to discovery requests, while excluding privileged materials.
2. To surface a critical subset of documents that require special attention, either because they support the client's case or pose risks if used by opposing counsel, so that appropriate responses can be prepared.

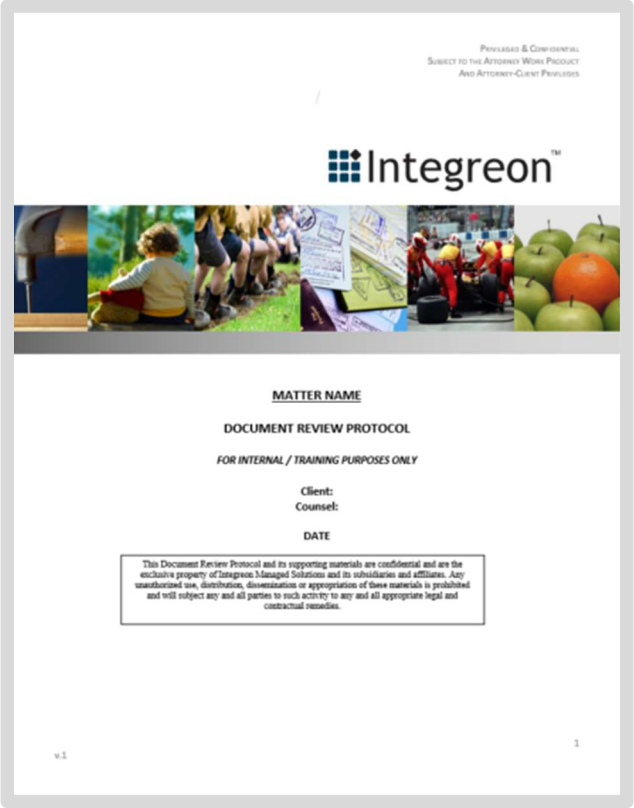


ii. Governing Documents

Achieving these goals requires structured planning and effective project management. During the planning phase, the project team should develop a core set of governing documents to guide both the execution and oversight of the review. These documents clearly define key decisions, workflows, and standards to ensure alignment across all team members.

Supporting Documentation should include:

- Document Review Protocol
- Project Management Manual or Standard operating procedure document with accompanying workflow diagram
- Query and Calibration logs



1. Document Review Protocol

Protocol templates provide standardized frameworks that can be tailored to the specifics of a given matter while maintaining a consistent structure and coverage. Once customized, the protocol serves as the review team's roadmap, outlining both the background of the matter and the procedures that guide day-to-day review decisions. The protocol should include:

- Context for the review effort, including litigation background and high-level objectives.
- Definitions and examples of relevance, responsiveness, and privilege.
- Confidentiality designations and guidance on their applications.
- Issue code guidance aligned with critical case themes.
- Treatment of special categories, such as documents with technical defects, foreign language content, and or varying levels of significance.

2. Project Management Manual or Standard Operating Procedure (SOP) with accompanying workflow diagram

The SOP

The implementation of repeatable processes, supported by comprehensive documentation, provides benefits that extend well beyond individual case outcomes. Teams operating under standardized procedures demonstrate greater efficiency—reviewers spend less time navigating procedural decisions and more time focused on substantive document analysis.

Standardization also enables more effective project management and resource allocation. With repeatable processes in place, project managers can better estimate timelines, assign resources appropriately, and identify potential bottlenecks before they impact deliverables. This predictability is especially critical when managing large-scale reviews under tight deadlines.

From a quality standpoint, repeatable processes create consistency that strengthens the defensibility of the entire review effort. If the adequacy or completeness of a review is challenged, teams can point to documented procedures, quality control measures, and systematic practices that demonstrate a good-faith effort to meet discovery obligations.

Comprehensive standard operating procedures translate high-level process designs into specific, actionable instructions that team members can follow consistently. SOPs should provide step-by-step

guidance for each process component, including examples and templates, and address common scenarios as well as exceptions.

Well-designed SOPs include:

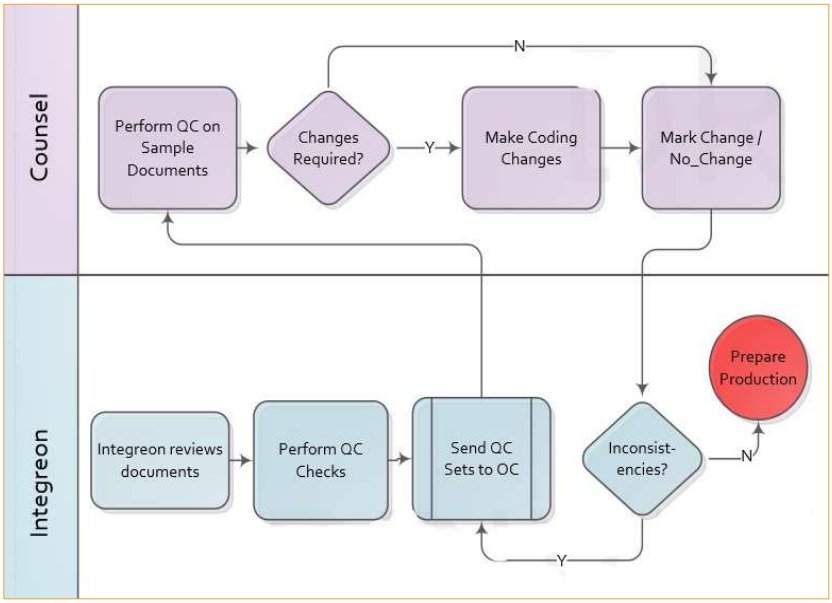
- Clear role definitions
- Specific quality standards
- Escalation procedures for atypical situations
- Mechanisms for regular review and updates

They should be detailed enough to ensure consistency, yet flexible enough to accommodate case-specific requirements and evolving circumstances.

This manual expands on the review protocol by establishing the operational framework for execution. It typically includes:

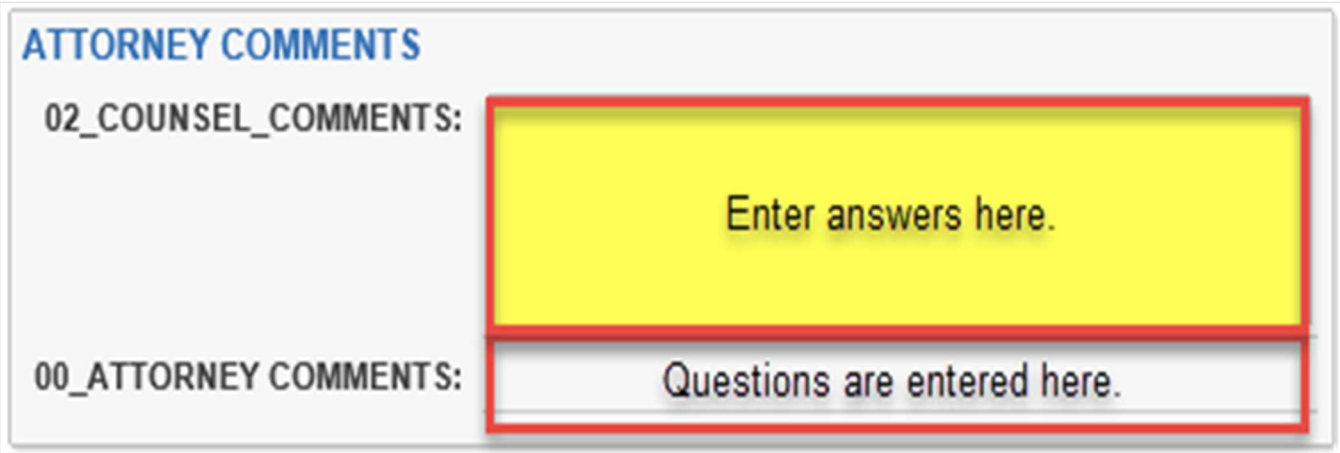
- Review scope and timeline, including key milestones
- Staffing structure, covering team roles, work schedules, and training plans
- Review workflow and platform configuration, ensuring seamless execution
- QC methodology, particularly for relevance and privilege determinations
- Communication protocols and feedback loops
- Reporting structure, including cadence and content
- Validation approaches, to confirm completeness and consistency

Workflow Diagrams



Visual workflow diagrams are essential tool for communicating complex processes to team members and stakeholders. These diagrams should clearly illustrate the flow of documents through each review phase, identify decision points and escalation procedures, and show the relationships between process components.

Effective workflow diagrams often use swim lane formats to show responsibilities across various organizational roles, decision trees for handling complex scenarios, and integration points where the hosting vendor, the document review vendor, counsel, and client intersect. These visual tools are especially valuable during team training, onboarding of new reviewers or stakeholders, and coordination of multi-vendor workflows.



3. Query and Calibration Logs

All coding decisions made by the review team should be clearly linked to documented guidance—either from the protocol itself, feedback provided during counsel's QC reviews, or entries in the query and calibration logs.

The Query Log plays a critical role in clarifying and expanding upon the guidelines outlined in the protocol. To maintain transparency and ensure consistent application, questions submitted through the query log should be inserted as comments on specific documents within the review platform. Counsel should provide corresponding answers or direction in a designated counsel comments box allowing for easy tracking and visibility.

These exchanges should be saved in a search within the review platform, accessible to both counsel and the review team. This shared resource helps ensure alignment and reduces repeated questions by making previous guidance easy to reference.

iii. Workflow Mechanics: Streamlining Review

Workflow design must address the review tagging structure, incorporating desired behaviors and constraints for individual tags. Counsel must consider how to handle document families, confidential or personal information, and redaction requirements. Data integrity and security protocols protect the review platform and process throughout the engagement.

1. Batching

Effective workflow design is both a science and an art. The way documents are presented to reviewers significantly impacts both the speed and quality of review. Batching strategies—such as applying filters to group documents by similar content, format, or custodian—help streamline workflow. Isolating discussion threads or grouping related documents allows reviewers to make consistent decisions more efficiently.

2. Coding Panels and Visual Clues

Review speed and accuracy improve when the review platform is configured with intuitive coding panels, keyword highlighting, and targeted visual cues. Highlighting specific search terms, flagging potentially privileged content, and organizing documents to focused sets helps reviewers navigate materials with greater precision.

iv. Document Review Workflow: Four Main Approaches

There are four primary approaches that form the foundation of effective document review workflow. Most successful review projects apply a thoughtful combination of these methods to balance efficiency, consistency and defensibility.

1. Linear Review

Linear review involves reviewing documents one by one. While this historically implied that they were reviewed sequentially, typically in the order they were collected or processed, modern linear workflows often incorporate prioritization strategies before documents are batched to reviewers, where possible.

Each document is examined individually, with coding decisions made on relevance/responsiveness, privilege, confidentiality, issues and document significance. The strength of this approach lies in its comprehensives and in the contextual understanding reviewers develop familiarity with the document population over time. However, it is time—and resource—intensive, and may delay the identification of critical documents until late in the review.

2. Continuous Active Learning (CAL), Technology-Assisted Review (TAR) and Predictive Coding

Machine learning-enhanced workflows like CAL, TAR, and predictive coding use algorithms to learn from human reviewer decisions and predict the relevance or characteristics of unreviewed documents. The system continuously improves its predictions based on ongoing reviewer feedback, allowing teams to prioritize the most likely relevant documents for human review.

CAL or TAR workflows typically begin with a training phase where reviewers code a seed set of documents, teaching the system to recognize patterns associated with relevant materials. As the system learns, it can rank the remaining document population by predicted relevance, enabling reviewers to focus their efforts on the highest-value documents first. This approach can either prioritize or significantly reduce the overall volume of documents requiring human review while maintaining high levels of accuracy in identifying key materials.

3. Single Instance and Propagation Review

Single instance review, also known as propagation review, focuses on eliminating redundant reviews of identical or near-identical documents. This approach identifies duplicate documents and the longest email thread, allowing reviewers to make coding decisions once and propagate those decisions across all instances of the same content.

This workflow is particularly valuable in modern document collections where email communications, shared documents, and copied files create substantial duplication. By reviewing unique content only once and applying those decisions systematically, teams can achieve significant efficiency gains while maintaining consistency in coding decisions. The approach requires sophisticated de-duplication and threading technology but can reduce review volumes by 30-70% in typical corporate environments.

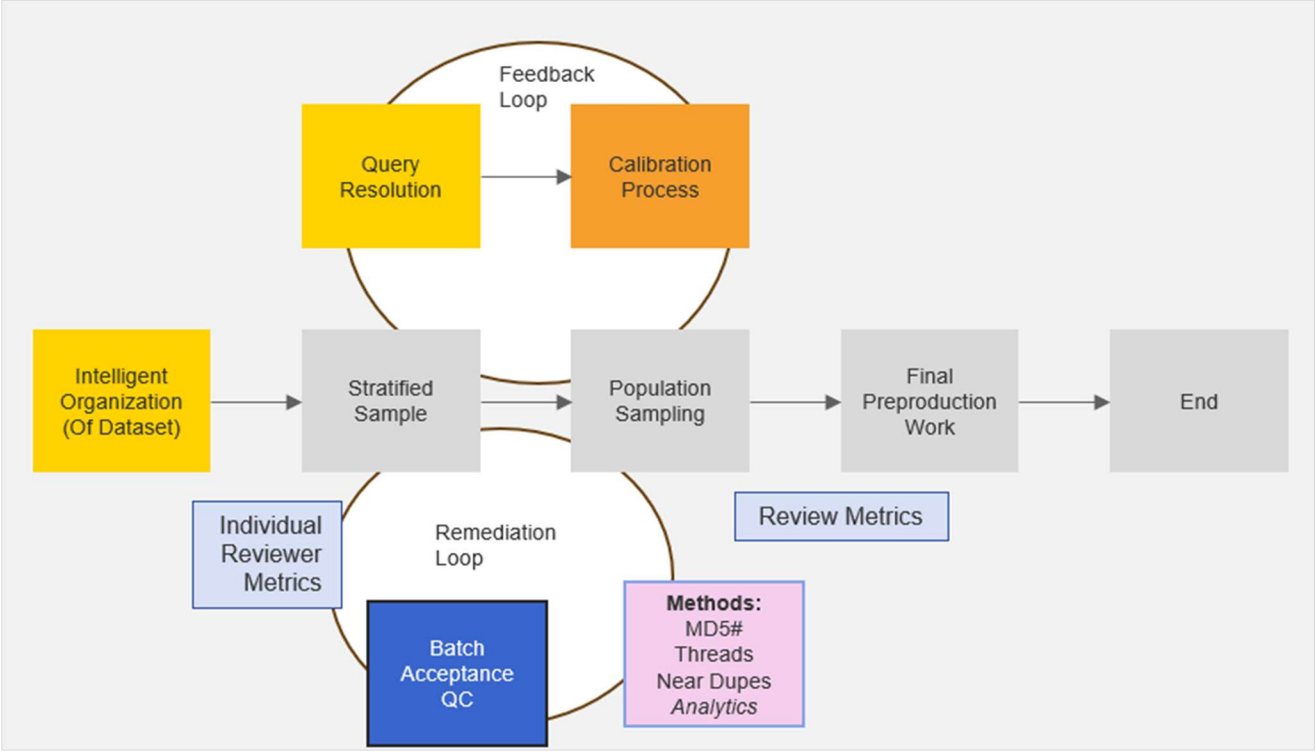
4. Generative AI Tools

The most recent addition to the document review toolkit involves generative artificial intelligence platforms like Relativity Air and similar tools. These tools analyze content to generate summaries, extract key facts, identify central themes and even draft initial coding recommendations.

Unlike traditional predictive coding that primarily focuses on relevance ranking, generative AI tools provides deeper, content-level insights—highlighting parties, dates, figures and factual context. This allows reviewers to focus on validation and nuanced decision-making rather than initial content analysis. The technology is particularly valuable for complex documents requiring detailed factual extraction or thematic analysis and is promising for post-production activities such as creating chronologies and deposition transcript analysis.

5. The Integrated Approach and Continuous Improvement

While each workflow method has its strengths, the most effective document review workflow typically relies on a thoughtful combination of multiple approaches based on the characteristics of the types of data the client corporation has. The key to maximizing both efficiency and accuracy lies in understanding when and how to deploy each approach for the workflow.



v. Quality Control

The review phase requires carefully structured workflows that define how documents flow through the system, who reviews what types of materials, and what escalation procedures exist for complex or uncertain documents. Repeatable review processes establish clear coding standards, define reviewer qualifications for different document types, and create checkpoints where senior reviewers or subject matter experts can validate decisions.

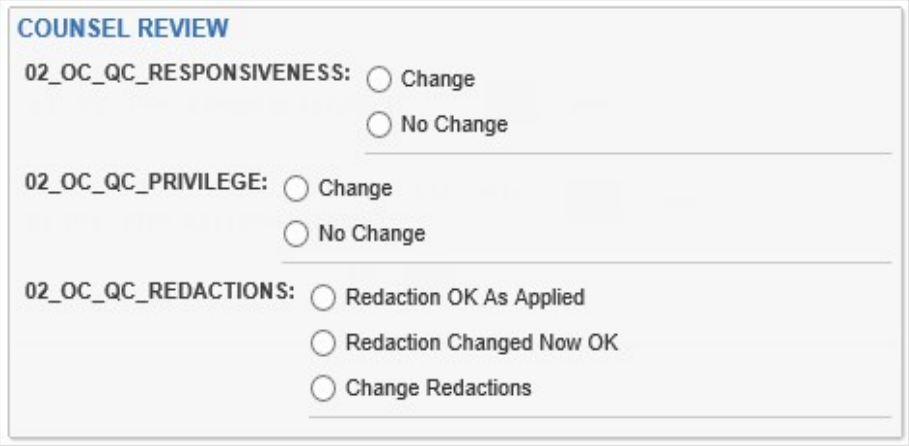
QC must be embedded throughout the review workflow—not just reserved for a final review stage. This includes statistical sampling protocols, reviewer batch QC, and continuous monitoring of coding patterns. When standardized and applied early, these QC processes provide real-time feedback, enabling course corrections before errors compound across large document populations.

Because human review is inherently imperfect, the goal is not perfection, but performance that reflects diligent effort and reasonable care. The distinction between inadvertent error and culpable negligence lies in whether counsel and the review team exercised reasonable care to avoid and detect mistakes.

Federal Rule of Evidence 502(b) illustrates this principle: inadvertent disclosure of privileged material won't result in waiver when the privilege holder "took reasonable steps to prevent disclosure" and "promptly took reasonable steps to rectify the error." This raises the crucial question: how does counsel define reasonable steps in document review?

Reasonable care means "defensible"—requiring intelligently designed quality control measures matched to rigorous training, performance measurement, and reporting. A robust quality control regime includes:

- Intelligent validation ensures the reviewable data set receives complete review by appropriate reviewers.
- Targeted review detects potential errors and identifies materials requiring further attention.
- Privilege isolation removes all privileged documents from production sets.



Counsel should focus quality controls on two critical areas: privilege designation and validation of presumptive production sets. Review results must be tested to ensure they remain consistent across the entire data set and team, multiple project phases, and protocol treatment for families and duplicates. Results must meet parameters for relevance, privilege, confidentiality, and issue coding while identifying all potential privilege.

Designing effective quality control presents significant challenges. When relying on sampling, counsel should employ statistical methods to identify representative random samples for re-review. The most defensible approaches combine sampling with targeted quality control elements that identify documents meriting second-level review while soliciting continuous counsel input to calibrate the team. All quality control elements should be documented with counsel's input.

vi. Redactions

Redaction processes demand particular attention to repeatability because of the irreversible nature of the work and the severe consequences of errors. Standardized redaction protocols must define what information requires protection, establish clear procedures for identifying and marking confidential information, and create validation workflows to ensure redactions are complete and appropriate.

Pharmaceutical eDiscovery often involves highly confidential data that must be protected before production. Standardized redaction protocols must define what information requires protection, establish clear procedures for identifying and marking confidential information, and create validation workflows to ensure redactions are complete and appropriate.

The process should include standardized redaction coding that distinguishes between different types of protected information, such as attorney-client privilege, work product, confidential business information, HIPAA or personally identifiable information. Each category requires specific handling procedures and different levels of review before finalization and includes:

- HIPAA and PII – Patient health records, employee data, and clinical trial participant information.

- Trade Secrets and Proprietary Data – Drug formulations, research findings, and IP-sensitive documents.
- Privileged Communications – Attorney-client discussions and work product.

When handling document reviews, particularly in pharmaceutical litigation or regulatory matters, it's crucial to distinguish between HIPAA-protected health information and PII to apply appropriate redactions.

1. HIPAA (Protected Health Information - PHI) Redactions

HIPAA covers Protected Health Information (PHI), which refers to any individually identifiable health data that is created, received, maintained, or transmitted by a covered entity (e.g., healthcare providers, insurers, business associates).

What Must Be Redacted Under HIPAA?

HIPAA's Privacy Rule mandates the removal of 18 identifiers when de-identifying PHI, including:

- Patient Names
- Addresses (smaller than state level, e.g., street address, ZIP code)
- Dates related to an individual (birthdate, admission date, discharge date, treatment date, death date)
- Phone numbers, fax numbers, email addresses
- Social Security Numbers (SSNs)
- Medical Record Numbers (MRNs)
- Health insurance or plan numbers
- Account numbers related to healthcare transactions
- License numbers (e.g., driver's license, DEA number)
- Vehicle identifiers, device identifiers, and serial numbers
- Biometric data (fingerprints, retinal scans, etc.)
- Full-face photographs and comparable images
- IP addresses and other electronic identifiers
- Any unique identifying characteristic or code

An example of a document that requires HIPAA redactions could be a doctor's note or adverse event reports that include a patient's name, diagnosis, treatment plan, and hospital admission date would require redacting all identifiers, ensuring the health information cannot be traced to an individual.

2. PII (Personally Identifiable Information) Redactions

PII refers to any information that can identify an individual, but it is broader than HIPAA and applies to both healthcare and non-healthcare settings. PII is regulated under various laws, such as the GDPR (EU), CCPA (California), and federal privacy laws beyond healthcare.

What Must Be Redacted Under PII Regulations?

- Full Name
- Home Address
- Phone Number
- Email Address
- Social Security Number (SSN)
- Driver's License Number
- Passport Number
- Bank Account Numbers, Credit Card Information
- Employment Information (e.g., company ID, job title, salary details)

Key Differences: HIPAA vs. PII Redactions

Feature	HIPAA or PHI Redaction	PII Redaction
Scope	Healthcare-related data	Any personally identifiable data
Regulations	HIPAA Privacy Rule	GDPR, CCPA, various U.S. privacy laws
Covered Entities	Healthcare providers, insurers, business associates	Any organization handling PII
Redacted Data	Health-related identifiers (medical records, treatment info) + standard PII	Names, SSNs, contact details, financial info
Example Use Case	Redacting patient details from medical records	Redacting personal details from contracts, HR files

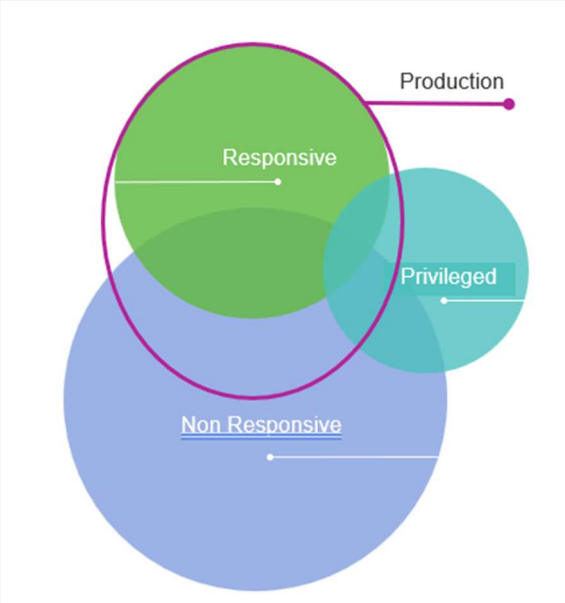
vii. Production

When producing documents to adversaries or requesting agencies, counsel and vendors must agree on production specifications—including format, metadata fields, and procedures. The processing and hosting vendor should supply comprehensive production logs that map production ID numbers (Bates numbers) to document ID numbers on the review platform and correlate to original data collection.

Document production represents the final stage the review process and requires systematic procedures to ensure that the final deliverable meets all technical and substantive requirements. Repeatable production processes include standardized metadata extraction, consistent file naming conventions, proper load file formatting, and comprehensive quality assurance testing.

Production workflows must also address the handling of native files, the conversion of documents to review formats, and the management of family relationships between documents. These technical requirements, when standardized, prevent the common production errors that can delay case resolution or compromise the usability of produced materials.

Once documents are reviewed and redacted, they must be produced in a manner that satisfies legal and regulatory standards while maintaining data security. Key considerations include:



- Regulatory Submission Compliance – Aligning production sets with requirements from agencies such as the FDA, EMA, and DOJ.
- Format Specifications – Delivering files in the appropriate format (e.g., TIFF, PDF, native in the load file).
- Bates Numbering and Document Tracking – Ensuring consistent labeling and accurate tracking of all produced materials.
- Secure Transfer Protocols – Using encrypted file transfers and access-controlled platforms to prevent unauthorized disclosures.

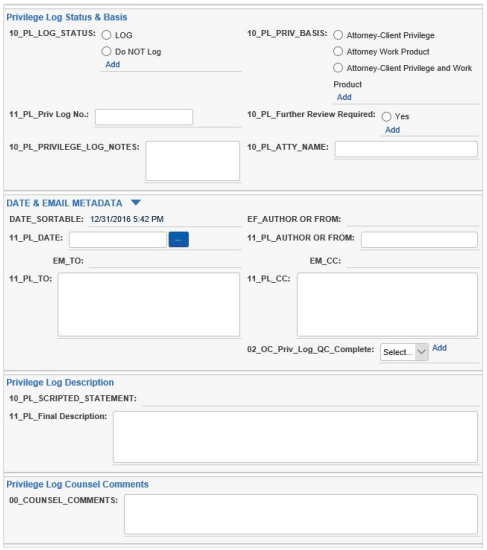
viii. Privilege Logs

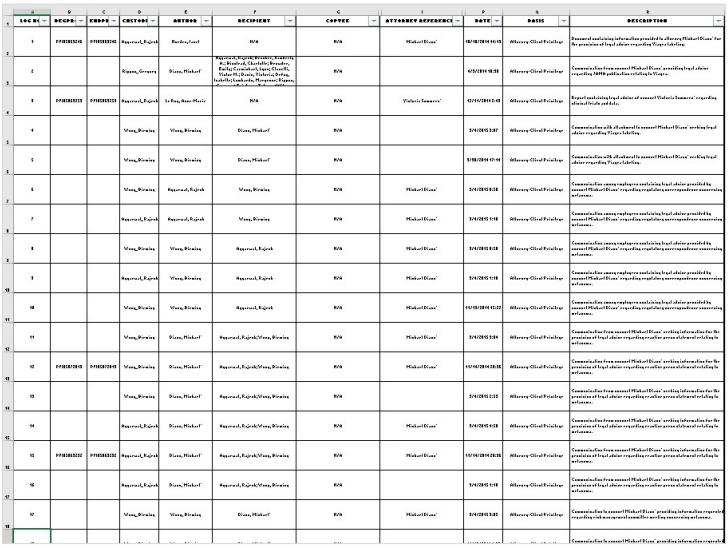
A privilege log is a critical component of the discovery process. It provides a line-by-line account of confidential communications, documents, and other materials that have been withheld or redacted due to claims of privilege. It's purpose is to identify these materials and substantiate the legal basis for withholding them. Following structured best practices ensures the log is both defensible and efficient.

Before drafting the log, consult the Stipulation or ESI Protocol and clarify the following questions:

- Should both fully withheld and partially redacted documents be logged?

- What metadata fields must be included?
- Is a categorical privilege log permitted?
- Should the format be in Excel, PDF, or something else?
- When is the log due? Is one required after each rolling production or only at the end?
- Are any documents excluded by date, subject matter, or sender/recipient (e.g., post-complaint documents or those involving specific counsel or topics)?





1. Organize the Population

Begin with the correct review set. Privilege logging should occur after documents have been reviewed for privilege—not during. Attempting to identify and log simultaneously leads to inefficiencies and inconsistencies.

If logging both withheld and redacted documents, keep them in separate sets. The criteria for evaluating and describing them differ, and separating them streamlines the process.

2. Think About the Outcome: What a Privilege Log Should Include

An effective privilege log is clear, concise, and defensible. A standard entry typically includes:

- Key metadata (Bates number, date, authors, recipients)
- Privilege basis (attorney-client privilege, work product)
- A description that conveys the general purpose of the communication without revealing privileged content
- Clear identification of legal counsel involved

3. Confirm Metadata Availability

Ensure you have all relevant metadata fields, including:

- Date
- Document type/extension
- Subject line or file name (if available)
- Author(s), recipient(s), including CC and BCC - normalized to remove extra content such as email addresses or extraneous formatting

Normalize and cross-check attorney names using a reference list, and use consistent identifiers (e.g., asterisk or "Esq.") to flag attorneys.

4. Draft Descriptions Efficiently with a Structured Workflow

A well-written description typically includes:

- Document type (e.g., email, memorandum, notes)
- Who is seeking or providing legal advice (e.g., attorney-to-client, client-to-attorney)
- The general subject matter of the communication, without revealing privileged content). To maintain consistency and efficiency, use predefined subject-matter categories such as:
    - Contract negotiations
    - Litigation strategy discussion
    - Draft agreements
    - Compliance issues

This standardization ensures descriptions are uniform, specific, and defensible.

5. Identify the Privilege Type

Use a picklist to consistently apply the privilege basis. The two most common options fall under attorney-client privilege and the work product doctrine; however, other types of privilege, such as common interest privilege (also called the joint defense privilege), may come into play for certain matters.

6. Create a Glossary for Certainty on Attorney Identification

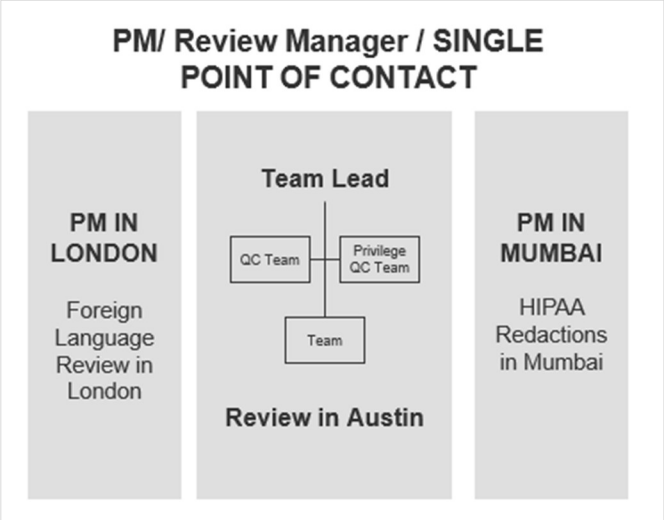
Accurately identifying attorneys and legal personnel is crucial. A glossary helps prevent errors such as mistaking non-attorneys for attorneys, and overlooking third parties whose involvement may waive privilege. At a minimum, your glossary should include:

- All outside counsel attorneys that are involved in the matter
- In-house counsel and their roles
- Associated law firms and legal teams
- Experts or third parties engaged for legal consultation

This ensures consistency across the log and reduced the risk of challenges from opposing counsel.

A well-structured privilege log is accurate, consistent, and defensible. By following these best practices—organizing the population, confirming metadata, standardizing descriptions, and maintaining a clear glossary—counsel can support privilege claims while minimizing disputes during discovery.

ix. Review Team Management, Selection and Training



The review manager serves as the single point of contact between the client, counsel, and the review team. This individual is responsible for overseeing day-to-day operations, managing communication, escalating questions, and ensuring alignment with counsel's directives. The review manager may be supported by a broader project team, including regional project managers (e.g., U.K.-based managers for non-English language reviews, India-based managers for redaction of heavy matters), as well as a team lead, a responsiveness QC team, and a privilege QC team. This layered structure ensures efficiency, specialization, and consistency throughout the review.

Selecting an effective review team begins with developing precise job descriptions and identifying the specific skills required for each engagement. Document review providers must establish clear protocols for recruiting, testing, and selecting team members. The complexity of the review determines the expertise level required. Simple redaction tasks for personal or confidential information require minimal legal training—paralegals working under attorney supervision can handle these effectively. More complex reviews demand seasoned judgment, making barred attorneys the optimal choice.

Every team member must receive comprehensive substantive training from counsel and thorough orientation to the selected review platform before starting work, which is usually done by the review manager. The review manager also proactively seeks counsel's guidance on any documents that the team has a question about. Counsel in turn confirms or corrects these early coding decisions. This iterative feedback loop ensures alignment with counsel's expectations and allows review protocols to evolve as additional guidance is provided.

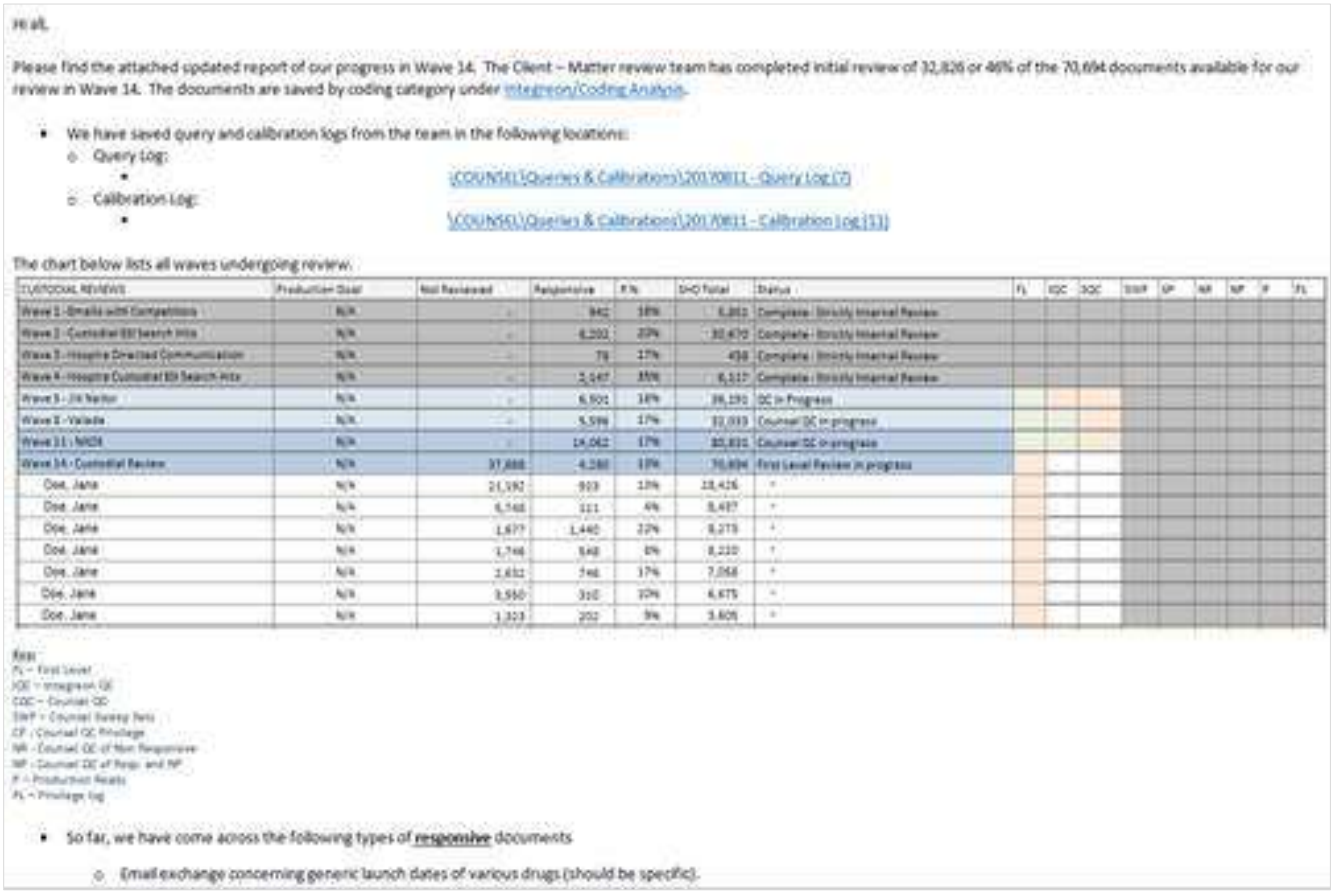
x. Communication and Status Reports

1. Communication

The Review Manager must establish formal, regular reporting and communication schedules among the review team, team leads, the hosting and data vendor, and supervising attorneys throughout the process. During ramp-up, counsel should remain available to confirm review guidelines and answer reviewer questions. Regular calls should be scheduled to review progress and address issues promptly.

The communication plan must document points of contact, escalation processes, and appropriate communication methods. This structure prevents miscommunication and ensures everyone stays aligned on objectives and expectations.

2. Status Reports



Hi all,

Please find the attached updated report of our progress in Wave 14. The Client – Matter review team has completed initial review of 32,826 or 46% of the 70,694 documents available for our review in Wave 14. The documents are saved by coding category under Integreon/Coding Analysis.

- We have saved query and calibration logs from the team in the following locations:
  - Query Log:
    - \COUNSEL\Queries & Calibrations\20170811 - Query Log (7)
  - Calibration Log:
    - \COUNSEL\Queries & Calibrations\20170811 - Calibration Log (11)

The chart below lists all waves undergoing review.

CUSTODIAL REVIEWS	Production Date	Not Reviewed	Responsive	R %	DHO Total	Status	FL	IQC	SQC	SWP	CP	NR	NP	P	PL
Wave 1 - Emails with Competitors	N/A	-	942	18%	6,261	Complete - Strictly Internal Review									
Wave 2 - Custodial ESI Search Hits	N/A	-	6,202	20%	32,470	Complete - Strictly Internal Review									
Wave 3 - Integrite Directed Communication	N/A	-	73	17%	438	Complete - Strictly Internal Review									
Wave 4 - Integrite Custodial ESI Search Hits	N/A	-	2,547	39%	6,517	Complete - Strictly Internal Review									
Wave 5 - JW Narbo	N/A	-	6,501	18%	36,191	QC in Progress									
Wave 8 - Valeda	N/A	-	5,596	17%	32,033	Counsel QC in progress									
Wave 11 - NIDA	N/A	-	14,062	17%	85,831	Counsel QC in progress									
Wave 14 - Custodial Review	N/A	37,868	4,280	10%	70,694	First Level Review in progress									
Doe, Jane	N/A	21,192	903	10%	18,426	"									
Doe, Jane	N/A	6,748	111	4%	8,437	"									
Doe, Jane	N/A	1,677	1,440	22%	8,273	"									
Doe, Jane	N/A	1,748	648	8%	8,210	"									
Doe, Jane	N/A	2,632	746	17%	7,058	"									
Doe, Jane	N/A	3,550	310	10%	6,675	"									
Doe, Jane	N/A	1,323	202	9%	3,605	"									

Key
FL – First Level
IQC – Integreon QC
CQC – Counsel QC
SWP – Counsel Sweep Pass
CP – Counsel QC Privilege
NR – Counsel QC of Non Responsive
NP – Counsel QC of Resp. and NP
P – Production Ready
PL – Privilege Log

- So far, we have come across the following types of responsive documents
  - Email exchange concerning generic launch dates of various drugs (should be specific).

Status Reports serve as the primary tool for presenting counsel with real-time information about review progress, designation breakdowns, and noteworthy documents. Review reports, generally issued every day when the review kicks off and as needed near the end of the review, deliver essential data on productivity, accuracy, operational issues, technical problems, QC sets released, and other requested metrics.

Counsel should read the status reports and proactively engage with the information they contain as status reports become meaningless without follow-through.

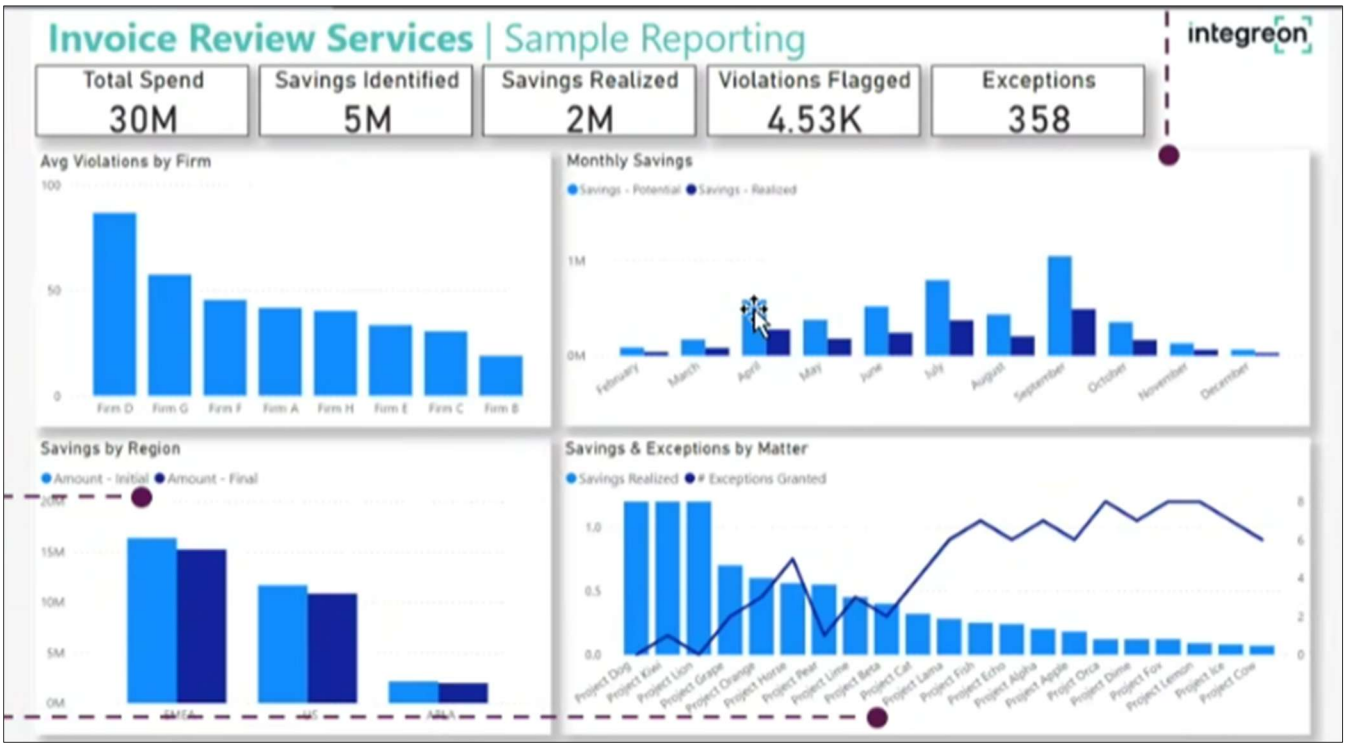
xi. Post-Case Documentation

Counsel and the client should determine early whether to maintain some or all documents in a repository for future or related litigation, then make necessary arrangements with the data and hosting vendor. Repositories offer significant advantages—once counsel makes final privilege designations, they can be preserved if the same dataset faces future or related litigation discovery.

As a final best practice, counsel and vendors involved in all discovery aspects must assemble complete documentary records of the discovery process, including collection, processing, review, and production specifications. This post-case documentation serves as a valuable historical record for answering later questions and demonstrating that counsel conducted discovery with diligence and reasonable care.

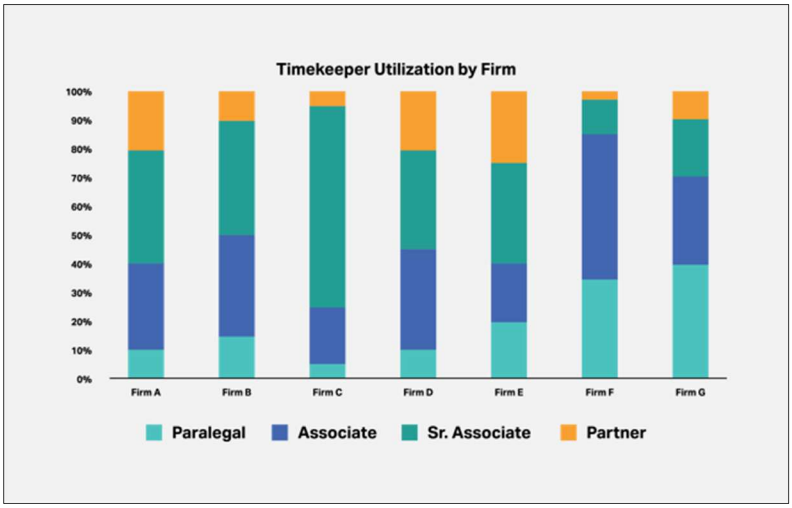
xii. Final Thoughts

Document review represents a critical, resource-intensive component of eDiscovery that demands active, competent project management following well-designed processes that reflect relevant best practices. Success delivers a timely, cost-effective, defensible work product that facilitates the overall litigation process and enhances favorable outcomes. The investment in proper planning, execution, and documentation pays dividends when the review withstands scrutiny and supports the client's interests effectively.



6. Tracking Work Using Task Codes and Legal Billing Framework

A robust system for tracking and managing eDiscovery and regulatory compliance tasks is essential for ensuring accountability, transparency, and defensibility. One key component is a task code system, which categorizes and tracks activities across departments. This allows for detailed monitoring, efficient reporting, and visibility into task progress. While an organization may opt to create a tracking and task code system bespoke to the specific needs of their in-house legal team, the Uniform Task Based Management System (UTBMS) is a good starting framework. This framework is maintained by the LEDES Oversight Committee and has been incorporated into most eBilling platforms.



Due to the high volume of discovery seen within the pharmaceutical industry, it is imperative that in-house teams have a clear understanding of who is performing what work and at what cost point for each stage of the discovery process. To achieve this level of granular insight into discovery spend, in-house teams should update existing outside counsel and provider billing guidelines to require use of the UTBMS L600 codes. This sub-set of the broader UTBMS framework focusses specifically on eDiscovery and breaks down various legal tasks into the core components of the EDRM (e.g., L612 for Legal Hold, L620 for Collections, L630 for Processing, L651 for Hosting, L653 for First Pass Document Review, and L655 for Privilege Review). Discovery workflow within the pharmaceutical industry often requires high volumes for redactions for sensitive PII, PHI, and HIPAA. The effort spent redacting sensitive content should be tracked under the L656 code and can then be reviewed across matters to establish a baseline of costs typically associated with this component of the discovery process.

Within this framework, a legal billing system is integrated to align task codes with specific activities—such as data collection, document review, and production—linking each to billable hours and associated costs. This ensures that every phase of the eDiscovery process is thoroughly documented and adheres to established legal billing standards. The result is a clear audit trail that supports internal oversight and external review, while enabling organizations to justify litigation or compliance-related expenses.

In addition to defensibility, this system improves resource management by revealing where time and costs are concentrated. Teams can make informed decisions about budgeting, staffing, and project

timelines. It also strengthens compliance by ensuring that all actions—from document handling to legal submissions—are systematically recorded and aligned with regulatory requirements.

V. Conclusion

As the pharmaceutical industry faces increasing regulatory scrutiny and litigation risks, a proactive and strategic approach to eDiscovery is essential. By implementing strong information governance policies, leveraging advanced technologies, and ensuring compliance with evolving legal frameworks, pharmaceutical companies can mitigate risks, enhance operational efficiency, and reduce eDiscovery costs. Organizations that adopt best practices for eDiscovery will be better positioned to handle regulatory inquiries, litigation, and internal investigations with confidence and precision.

VI. About the Authors

Phoebe Gebre

Vice President
Litigation Services

Based in New York, Phoebe is a subject matter expert in Managed Document Review and has been with Integreon since 2010. Her expertise extends to privilege logs, data visualization, pre-trial support, and document review workflows that incorporate machine learning and analytics. Phoebe works closely with clients to find efficient and cost-effective solutions to litigation issues while building long-term trusted relationships.

Phoebe has over a decade of experience working with multinational corporations across various industries including pharmaceuticals, technology, and financial services. She has supported project teams in multiple global locations and was integral to opening Integreon's Austin delivery center successfully implementing an infrastructure and people system which earned her an Outstanding Service Award from a client. Her inquisitive approach creates a space for deep analysis, thoughtful innovation, and collaboration. Her strength is distilling complicated issues into simple, actionable items to align multiple stakeholders for maximum efficiency.

Phoebe received her B.A. in Philosophy and her BBA in Computer Information Systems from Georgia State University and her J.D. from Rutgers School of Law - Newark. She is a member of the New York bar.

Scott Bien

Vice President
Litigation Services

Scott Bien is the Vice President of Enterprise Solutions at Integreon. Based in Cincinnati, Ohio, Scott leads strategic initiatives that help corporate and law firm clients optimize legal operations, automate processes, and integrate advanced technologies into their workflows. He also serves on Integreon's Innovation Committee, where he contributes to evaluating and deploying emerging generative AI applications across legal services.

Prior to joining Integreon in 2023, Scott held several leadership roles at UnitedLex, including Vice President, Enterprise Data and Insights, Vice President, Cyber Incident Response, and Director, Digital Contracting Solutions.

Scott holds a J.D. from Northern Kentucky University's Salmon P. Chase College of Law and a B.A. in History from the University of Cincinnati. He is a licensed attorney with the Ohio Bar.

Robert Daniel

Senior Director
Litigation Services

Robert joined Integreon in 2017 as Senior Director. Leveraging his many years of practical and tactical experience solving complex discovery related challenges, Robert is an SME for eDiscovery as well as Financial Services clients. Given the many complex solutions he helped to both create and implement in the past, this knowledge base is available to all clients of every size. Robert has also helped to co-create and run a new service line at Integreon focusing on Subpoena Processing Compliance services. Integreon has reacted to client needs to address this growing challenge for many companies in financial services, insurance, telecom, and internet service provider industries.

Prior to joining Integreon, Robert worked for Bank of America for 25 years. Leadership roles included Senior Executive Manager of the Legal Discovery Operations team, responsible for assisting internal and external counsel in the identification, collection, processing and review of data to comply with regulatory and court demands, and Director of Regulatory Investigations within the Litigation department, responsible for responding to high profile/high risk SEC, OCC, FRB, FINRA, CFTC, and Financial Crisis Inquiry Commission inquiries for the corporation.

Robert is also engaged in two industry working groups focused on the rewrite of the Information Governance Reference Model (IGRM) and the Collection Phase rewrite of the Electronic Discovery Reference Model (EDRM).

Claire Frazer

Managing Director
Review Solutions
Product Management

Based in the United Kingdom, Claire joined Integreon in 2013 and is the product manager for review solutions. In this role, Claire applies her deep expertise and client centric approach to further evolve Integreon document review services. This includes developing new offerings to help clients meet their long-term goals, identifying enabling technologies and workflows, cultivating strategic partnerships, and ultimately ensuring the highest standards for quality client delivery.

Claire has been working in the legal industry since graduating in 2004. She started her career as a paralegal in private practice before moving into the charity sector and, in 2008, Claire went on to be admitted to practice law in England and Wales. Prior to joining Integreon, Claire gained extensive experience in several disciplines, including litigation.

Before redirecting her focus to product management, Claire was responsible for overseeing all document review delivery operations globally. Prior to taking on responsibility for delivery operations, as a project manager, Claire managed a wide range of projects involving complex litigation, regulatory proceedings, and government investigations, for an array of companies and industries, including Fortune 500 clients, large financial institutions and some of the world's top ranking law firms. Claire's experience also encompasses many different types of review workflows from basic linear review to complex, analytics driven, workflow solutions.

Miguel Villalobos

Sr. Director -
AI & eDiscovery

Miguel joined Integreon in 2024 as a Senior Director, AI and eDiscovery. In this role, Miguel is responsible for assessing and evaluating the eDiscovery technology landscape including the application of gen AI. He is a frequent thought leadership contributor both publishing and participating at industry seminars and webinars. Additionally, Miguel works closely with Integreon litigation clients to identify best-fit technology solutions to streamline processes and reduce costs.

Miguel has over 12 years of experience applying advanced technology to eDiscovery to reduce the time and cost of litigation preparation while delivering accurate results. He has implemented machine learning and AI technology solutions reducing client litigation spend in large and small matters by over 73% in the life sciences, financial services, and agricultural sectors. Miguel has a JD from The Ohio University Moritz College of Law, and bachelor's degrees in philosophy and English from The Ohio State University.