Legal Data Intelligence

Internal Investigations

Updated: May 5, 2025

Inquiries conducted within an organization to examine allegations of misconduct, compliance violations, or other issues.

USE CASE IN PRACTICE

Regulatory Inquiry Response Strategy

The compliance team at a pharmaceutical company initiated an internal investigation following a regulatory inquiry. Given the unprecedented size and scope of the data that required analysis, the company's Legal Data Intelligence practitioner leveraged the LDI model to identify the SUN (sensitive, useful, necessary) data within both internal and external communications. This approach enabled the compliance team to provide a comprehensive and thoughtful response to the regulator.

Targeted Discovery in Sensitive Investigations

A company tasked its Legal Data Intelligence leader with helping investigate alleged workplace harassment. Leveraging the LDI model, the leader identified and analyzed relevant employee communications in which employees were referring to the individual involved. Filtering ROT (redundant, obsolete, trivial) data from SUN (sensitive, useful, necessary) data, the LDI practitioner narrowed the dataset to 500 key documents, enabling the firm to make informed decisions on next steps.

Proactive Litigation Readiness

A utility company conducted an internal investigation in anticipation of a class action lawsuit over its role in a major wildfire in California.

MODEL WORKFLOW

Initiate

Scope Project

Set protocols for the investigation to ensure it is in line with internal policies and procedures, as well as legal requirements

Consider if there is a case to answer

Assemble an appropriate team of subject matter and technical experts and identify if there is specific external expertise that may be required (e.g., external and internal counsel, business leads, IT leads, LDI practitioner, forensic expert)

Identify risk of litigation or potential regulatory action depending on result of investigation

How Technology Can Assist

Enables analysis of the merits of the case and comparisons to similar or historic cases

Captures and organizes objective metrics for each matter (e.g. fees, costs incurred, time to resolution) to enable informed decision-making

Identify Data

Define data sources, custodians, and specific criteria (e.g., date ranges) based on information available

Determine if custodial interviews and discussions with IT administrators regarding messaging, file storage, and other relevant data sources are warranted/proportionate

Ensure identification exercise is sufficiently robust and in line with proportionality (consider if you need an expert to assist) and periodically reevaluate to account for additional sources discovered

Consider alternative and additional data sources and approaches (e.g., social media analysis)

How Technology Can Assist

Tracks data sources from identification through preservation, collection, and processing to ensure nothing is missed

Assists with the identification of relevant data sources and locations

Analyzes common data to identify additional data sources or off-channel communications

Notes

Ensuring sufficient identification may be a particular consideration if there is risk of related litigation or regulatory action pending the result.

Determine if a client's in-house systems have search and identification functionality that is sufficiently robust.

Preserve and Collect Data

Consider whether data can be preserved in place by use of robust legal hold or similar functions, or whether data should be collected to preserve it

Determine if covert data collection is appropriate and plan accordingly

Consider if data needs to be collected in a forensic manner (e.g., if there is potential risk of resultant civil or criminal litigation or regulatory action)

How Technology Can Assist

Identifies potential data sources instead of conducting interviews

Provides legal hold functionality to preserve data in place and ensure it cannot be deleted (where appropriate)

Collects data covertly, remotely, or both in many cases

Tracks the implementation and oversight of each legal hold, including periodic reminders and audits to ensure compliance

Ensures data that should be preserved is correctly secured as required (note that once preserved, data need not be processed until or unless needed)

Ensures data is preserved and collected while maintaining both file and system application and system metadata (if appropriate)

Notes

More challenging sources, or sources containing potentially ephemeral data like mobile devices, may need to be imaged swiftly to ensure preservation.

Extract and Process Data

Select and extract the data to be processed from the data that has been collected so it can be loaded onto a review and production platform

How Technology Can Assist

Applies filters and other criteria to minimize data to be extracted and processed, thereby reducing the processing of ROT data

Quickly manages large data volumes

Reads and records metadata of files for searching and analytics

Extracts attachments and expands compound files

Indexes content to enable searching and analytics

Identifies encrypted files and files that could not be read for exception handling

Eliminates manual workflows and reduces human error

Reduces hosting of ROT data and thus hosting cost through de-duplication, de-NISTing, etc.

Automates the preservation of electronically stored data (mailbox and laptop data, network drives, SharePoint, etc.)

Pulls data from sources into a platform for processing

Reduces collection of ROT data by excluding non-relevant documents by using date filters, export rules to filter junk/system files, etc.

Investigate

Search and Evaluate Results

Run searches/analytics/machine learning to find pertinent data, review the search results, and cull unresponsive data

How Technology Can Assist

Allows for complex Boolean searches and filtering using content and metadata

Performs complex searches and summarizes large data sets using generative AI

Enables custom tagging and foldering to classify and stratify data

Reduces review volume/ROT data with tools like email threading, deduplication or suppression of near-duplicates, categorization, and clustering

Codes documents based upon training or prompts using traditional and generative AI tools

Creates custom chronologies using traditional and generative AI tools

Notes

Use of analytics and AI should be subject to local regulation, transparency requirements, and ethical constraints.

Analyze Data

Examine data to find relevant information, privileged information, personal information, patterns, and insights

How Technology Can Assist

Performs complex searches, including regular expressions

Identifies specific data sets and privileged documents through prompts and training using generative AI

Conducts visualization methodologies and gap analysis

Summarizes files for potential use in listings (e.g. privilege logs)

Provides explanations and citations to help validate output

Codes information for privilege/privacy, which is useful if it's determined that a chance of litigation could come later

Analyzes emails/files for signs of manipulation or fabrication

Enables digital forensic analysis to identify and/or recover deleted material or extract non-user information that may be relevant (e.g., web history information, recently connected devices)

Notes

Use of analytics and AI should be subject to local regulation, transparency requirements, and ethical constraints.

Apply Strategic Decision-Making

Determine the information relevant to the investigation

How Technology Can Assist

Generates summaries of documents or groups of documents using generative AI

Generate timelines of actions through prompts and training using generative AI

Identifies key questions that should be asked of an employee based on findings in the documents and the allegations that have been made using generative AI

Notes

Use of analytics and AI should be subject to local regulation, transparency requirements, and ethical constraints.

Implement

Act on Results

Determine next steps based on findings

Record/Document

Document results and metrics for use in future analytics and analysis

Document any lessons learned to assist with future matters