

The Application of Legal Data Intelligence to Governmental Investigations

By: Kevin M. Clark, Melina Efstathiou, Chuck Kellner, Daniel Miller, and Kelly Friedman



Regulatory investigations are commonplace in today's global economy, ensuring compliance and trust across industries. In fact, 70% of organizations were involved in at least one regulatory proceeding in 2024, up from 61% in 2023. As scrutiny intensifies from regulators, organizations must respond swiftly to mitigate risks and gain strategic advantages.

Leveraging advanced technologies and streamlined workflows enhances credibility, manages costs, and ensures timely compliance, fostering transparency and accountability in an ever-evolving regulatory landscape. In this paper, the authors explore the use of Legal Data Intelligence in the investigations context, drawing on experience in both the United States and the United Kingdom.

To better understand how technology is transforming this space, it is essential to define the foundational concept that underpins many of these advancements: Legal Data Intelligence.

What is Legal Data Intelligence?

Legal Data Intelligence (LDI) refers to data comprehension gained with structured use of advanced analytics, machine learning, and legal technologies to extract meaningful insights from complex, often fragmented, datasets. In the context of litigation, investigations, and regulatory response, LDI enables legal teams to move beyond reactive discovery and toward proactive legal strategy—leveraging data patterns, stakeholder mapping, risk profiling, and predictive modelling to improve decision-making under pressure.

To illustrate the impact of LDI, we first consider the regulatory and investigative landscape in the United States and then consider LDI in the context of Public Inquiries and Inquests in the United Kingdom.

Regulatory Investigations in the US

Regulatory enforcement in the United States reached a record \$8.2 billion in penalties in 2024, driven by cases in cryptocurrency and AI. Swift responses and efficient workflows are crucial in handling regulatory requests.

Regulatory scrutiny on technology, privacy, data breaches, consumer lending, and employee relations affects all regulated or publicly traded companies. Emerging tools that are shaping the future of investigations include generative AI for narrative summarization, flagging inconsistencies between witness statements and exhibits, real-time entity resolution across data types, and predictive modelling to assess likely points of regulatory concern. Platforms now offer dynamic visualization of custodian communications, anomaly detection in structured datasets, pre-identification of privilege, and automated redaction tools that reduce review bottlenecks and human error.

Traditional workflows for collecting and reviewing data are simply too slow and inefficient for today's regulatory environment. Consider instead an LDI workflow with the following elements:



This kind of workflow allows a legal team to increase credibility with regulators, essential for resolving regulatory investigations effectively. Key stakeholders are satisfied with the swift and effective handling of the investigation. Collaborative storytelling and automated timelines help present a clear narrative to the government. This approach ensures a strategic and tactical advantage in regulatory investigations by leveraging technology for speed and efficiency.

For example, during a recent Foreign Corrupt Practices Act investigation, a global company deployed LDI techniques to identify whether payments, gifts, or other benefits were offered, promised, or given to foreign officials to influence decisions or gain an unfair business advantage. By leveraging advanced data analytics and targeted pattern recognition, the company was able to reduce the dataset by 50% before review even began. As a result, the overall document review workload was reduced by an estimated 70%, significantly cutting time and cost while increasing focus and precision. This strategic use of LDI not only streamlined the investigation process but also allowed legal teams to present regulators with a clearer, data-backed narrative.



Public Inquiries and Public Inquests in the UK

In the UK, the challenges with disclosure in Public Inquiries and Public Inquests are numerous and in direct conflict with strict timelines and budget restrictions. The graphic below highlights key challenges:



The mass disclosure of documents can lead to an Inquiry receiving a huge volume of data to be controlled and processed, in direct conflict with budget restrictions, time constraints, and resource limitations. Documents are often disclosed in different formats, making it difficult to process and analyze the information effectively. Multiple documents are often disclosed from different sources leading to the time-consuming task of removing duplicates. Near-duplicates add an additional challenge. Further, varying levels of redaction must be applied to documents which are disclosed to the Inquiry. The task of redacting personal sensitive information, irrelevant, and privileged information can be time-consuming and monotonous, but must be done correctly to comply with General Data Protection Regulation (GDPR) privacy obligations. A procedural practice known as Maxwellisation, to ensure fairness to individuals who may be criticized in the final report, further complicates matters. All of these challenges require meticulous planning and robust procedures to ensure that the inquiry can proceed effectively and transparently.

The following are the key stages to an LDI approach to the challenges:



By meticulously addressing these stages, organizations can navigate complex legal landscapes with confidence and precision, ensuring the project's success and compliance with relevant legal standards.

An LDI practitioner plays a crucial role in the identification of the right stakeholders and variables of different systems. By leveraging advanced data analytics and machine learning algorithms, legal teams can systematically analyze vast datasets to pinpoint key influencers and decision-makers whose expertise will be invaluable for the project's progression. This process involves scrutinizing historical data, previous case outcomes, and stakeholder engagement patterns to make informed selections.

Moreover, LDI teams can assist to accurately define the variables within systems that will significantly impact the project's objectives. Through sophisticated data modelling and simulation, legal teams can identify critical data points, potential bottlenecks, and access constraints that might affect the legal compliance and overall success of technology deployments. This intelligence-driven approach ensures that all relevant variables are considered, mitigating risks and enhancing the precision of legal strategies.

In a recent high-profile Public Inquiry in the UK, a law firm applied LDI techniques to generate lists of key characters and chronologies to bring priority entities and events to counsel's attention. Furthermore, as a retrospective measure, legal professionals applied generative AI features to compare public information to their own findings. As a result, the work was conducted in an efficient way, tackling the time and cost restrictions while allowing the legal professionals to avoid scrutiny by quality checking their own previously applied work product against the predictions of an intelligent system.



Key Considerations for an LDI-focused Approach

Below, we provide an outline of the considerations when applying an LDI approach to governmental investigations.

A. Scope the Project

1. Case Parameters

- **Objective:** Determine the ask of the Investigation/Inquiry.
- **Scope:** Outline what will be included and excluded from consideration, with respect to custodians, date ranges, and issues.
- Stakeholders: Identify all parties involved or affected by the Investigation/Inquiry.
- **Constraints:** List any limitations or restrictions that might impact the delivery of work product, such as budget, time, resources, complexity of data, and accessibility of systems.

2. Deliverables with Timelines

- **Milestones:** Break down the management of the case into key phases or milestones. Each milestone should have a clear deliverable.
- Timeline: Create a detailed timeline for each milestone with associated dependencies.
- **Dependencies:** Identify tasks that depend on the completion of other tasks. This helps in understanding the sequence of activities and managing potential bottlenecks.
- **Review Points:** Schedule regular review points to assess progress and make necessary adjustments.

3. Data to Fall Within Scope

- Data Requirements: Specify the types of data which are in scope.
- **Data Sources:** Identify where the data will come from. This could be internal databases, external sources, or a combination of both. Be sure to consider active, as well as archived data.
- **Data Collection Methods:** Outline how the data will be collected. This requires proactive advice and guidance from forensically trained individuals.
- **Data Quality:** Establish criteria for data quality to ensure the data is accurate, reliable, and relevant.
- **Data Security:** Ensure that data handling complies with relevant data protection regulations and standards.

B. Identify Stakeholders

1. Types of Stakeholders

- **Internal Stakeholders:** These are individuals or groups within the organization who should be directly involved in the case. Examples include:
 - Legal teams



- o Technology teams
- Data security teams
- **External Stakeholders:** These are individuals or groups external to the company under investigation that need to be included in the case. Examples include:
 - o Legal teams
 - o Counsel
 - Experts
 - Technology teams
 - IT teams
 - Forensic technology experts

2. Selecting Appropriate Stakeholders

Expertise and Knowledge: Choose stakeholders who possess the necessary expertise and knowledge relevant to the case. This ensures informed decision-making and valuable insights.

3. Engaging Stakeholders

- **Roles and Responsibilities:** Clearly define the roles and responsibilities of each stakeholder. This helps in setting expectations and ensuring accountability.
- **Communication Plan:** Develop a communication plan to keep stakeholders informed and engaged throughout the project. This includes regular updates, meetings, and feedback sessions.
- **Feedback Mechanisms:** Establish mechanisms for stakeholders to provide feedback and input. This can be through surveys, focus groups, or one-on-one meetings.
- **Conflict Resolution:** Have a plan in place to address any conflicts or disagreements among stakeholders. This ensures smooth collaboration and progress.

C. Determine Data in Scope

Data Requirements: Clearly define what data would fall within the scope of this Investigation/Inquiry. This involves understanding the case's objectives and determining the specific data points that will support these goals.

- Data Types: Identify types of data.
- Data Sources: Determine where the data will come from.
- **Data Relevance:** Ensure the data collected is relevant to the scope of the Investigation/Inquiry. Irrelevant data can lead to unnecessary complexity and confusion.
- **Data Quality:** Establish criteria for data quality and understand issues with any specific data sources (e.g., hard copy documents).
- **Data Volumes:** Understand the overall data volume and how it varies by source, as some data types (e.g., video) are especially large.

D. Establish System Accessibility

- Access Parameters: Consider how system access will be controlled and determine user roles and permissions.
- **Data Security:** Encrypt data at rest and in transit, use secure file transfer methods, ensure proper data segregation, and maintain a clear, documented chain of custody and audit trails.

E. Set Priorities

- **Criteria:** Consider factors such as urgency, importance, accessibility, dependencies, and resource availability.
- **Usage:** Create priority lists and clearly communicate priorities; conduct ongoing reviews and make adjustments as required.

Conclusion

As regulatory expectations evolve and the volume, complexity, and sensitivity of data continue to grow, legal teams are under increasing pressure to respond with speed, accuracy, and strategic foresight. LDI is not a futuristic concept, it is a present-day necessity that enables legal professionals to move beyond reactive data processing and toward proactive, intelligence-driven decision-making. From US regulatory investigations to UK Public Inquiries and Inquests, LDI offers a consistent framework for navigating fragmented data ecosystems, aligning stakeholders, and managing disclosure obligations with greater efficiency and control. By integrating advanced analytics, machine learning and generative AI, and well-scoped project planning into legal workflows, organizations can reduce cost, mitigate risk, and present more coherent, compelling narratives to regulators and tribunals alike.

Ultimately, the legal function is undergoing a transformation—from one centered on documentation and defense to one empowered by data, insight, and influence. Those who embrace LDI now will be best positioned to lead, not just comply, in this next era of legal complexity.

