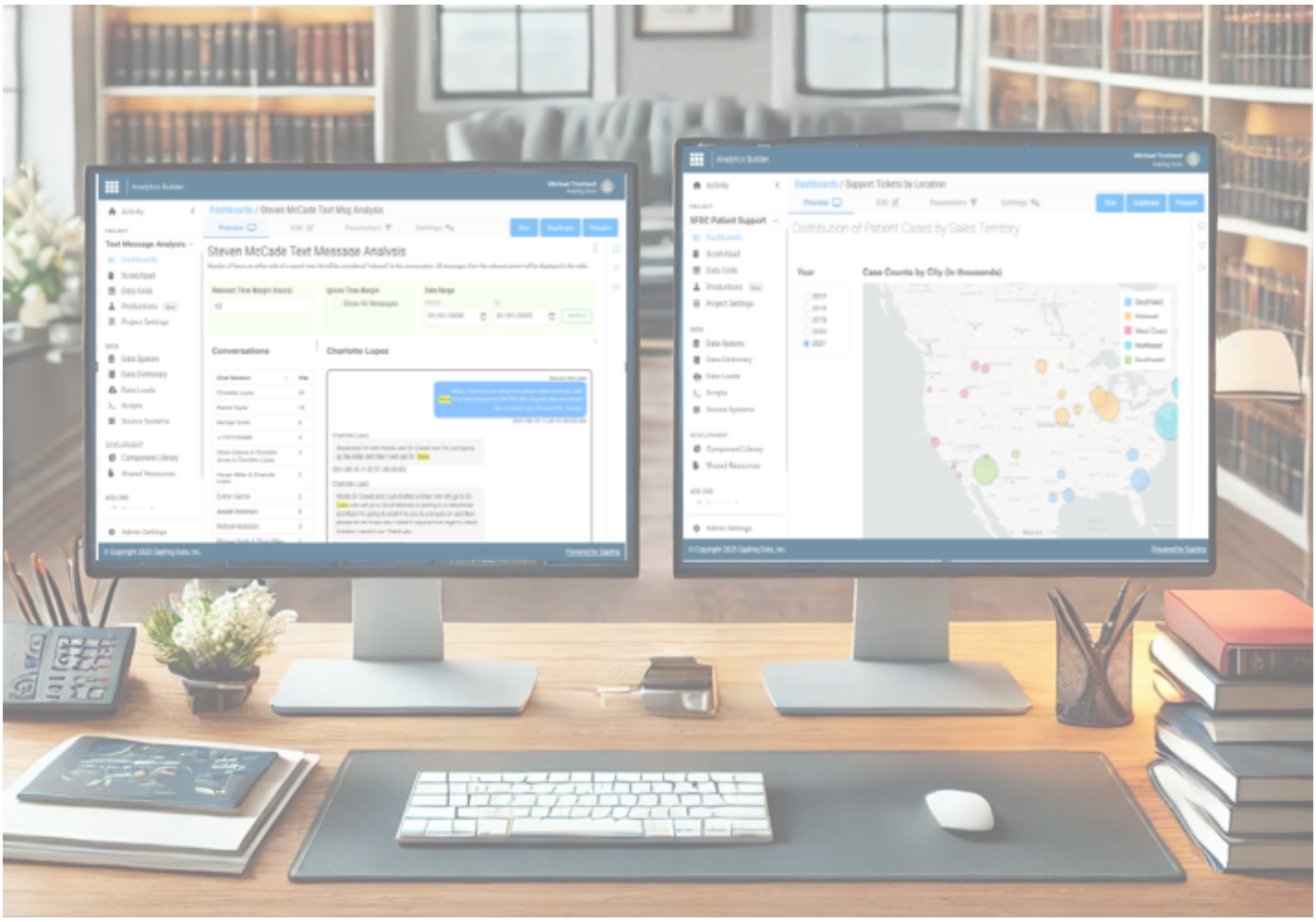
sapling

Data, Not Documents

Unlocking the True Value of Your Digital Evidence



Data, Not Documents

Unlocking the True Value of Your Digital Evidence

Introduction

When it comes to litigation and other legal proceedings, the nature of evidence has undergone a seismic shift over the past few decades. A world that was once dominated by paper-based records, physical files, and manually stored information has now become a digital landscape overflowing with emails, chat messages, social media posts, cloud-based collaboration data, and dynamic structured and unstructured data sources. This evolution has profoundly impacted how discovery is conducted – spawning an entirely new industry called "eDiscovery" – forcing legal teams to rethink how they manage and analyze evidence in litigation and investigations.

Despite this transformation of evidence to digital data, the legal industry has continued to operate under a document-centric paradigm and looks to treat everything as a "document" within the eDiscovery life cycle. For the most common sources of evidence over the past two decades – emails and office files (like Word documents and PowerPoint presentations) – treating them as "documents" (and converting them into TIFF or PDF files for production) wasn't much of a stretch as these file formats were very document centric.

However, not all sources of electronically stored information (ESI) during that time fit neatly into the document paradigm over that time. For example, legal teams used to try to convert Excel files into a static image format like TIFF files for production, but doing so was problematic, as Excel files are often not formatted for printing and conversion to TIFF stripped out key metadata, like the formulas behind the displayed numbers. Today, most discovery teams have standardized on producing Excel files "as is" to eliminate these challenges.

"I thought this was going to be a document review??"

As ESI has continued to evolve, it has continued to resemble documents less and less. Text and chat messages, for example, are short forms of communication that are conversation oriented, not document oriented. Yet, many legal teams have taken these messages and converted them into static documents with messages grouped by an arbitrary time frame (typically, 24 hours or the entire conversation) – before even analyzing and reviewing them. The move to the cloud has increased the use of enterprise solutions (like Salesforce), which are database oriented, not document oriented. **Text/chat messages and enterprise solution databases are just two examples of how evidence today is available for discovery as data, not documents.**

Yet, legal professionals continue to apply the document mindset to evidence, even when that evidence doesn't begin to resemble a document format. Want one example of that? Most people in the legal industry call the Review phase of discovery "Document Review", regardless of the format of the evidence being

reviewed. This document mindset limits visibility into the true nature of digital evidence, obscuring critical relationships, context, and patterns that exist beyond individual files. As a result, legal teams often struggle to extract meaningful insights, leading to inefficiencies, increased costs, and missed opportunities to strengthen their case strategies.

To fully harness the power of digital evidence, legal teams must start to think of the evidence as data, not documents. A data-driven approach treats digital evidence in its native form, preserving its structure, interactivity, and (very importantly) metadata. Taking a "data, not documents" approach enables more advanced analysis, including visualization of communication patterns, automated detection of anomalies, and AI-driven insights that are currently unavailable when legal professionals rush to force the evidence into their document paradigm. **By embracing a "data, not documents" approach to evidence early in the case, legal professionals can unlock the full evidentiary value of their digital assets, improving efficiency, accuracy, and decision-making in the process.**

This white paper explores the evolution of evidence from paper to digital, the limitations imposed by a document-centric approach to eDiscovery, and how a "data, not documents" approach can transform your eDiscovery workflows. You may never look at digital evidence the same way again!

How We Got Here

So, why is the legal profession so document-centric? And what's driving a need to start thinking about the evidence as data, not documents? Let's take a look at how we got here and where we go from here.

Why the Legal Profession is So Document-Centric

For centuries, the legal profession has been inextricably linked to documents. From handwritten contracts and bound case law volumes to typewritten pleadings and printed evidence exhibits, legal work has historically revolved around the creation, exchange, and interpretation of documents. In a paper-based world, this made sense—documents were the primary means of recording agreements, preserving evidence, and presenting information in court. The legal system itself was built on the assumption that evidence would exist in neatly contained, tangible formats that could be reviewed page by page.

As we entered the digital age, physical paper documents first began to evolve into images of those documents, which led to organizations implementing huge scanning facilities to create image copies of paper documents. As a result, most electronic evidence at first was simply an electronic representation of the physical paper documents, which even included the staple marks and handwritten notes! Of course, a picture of a document provided no automatic evidentiary value, which is why the process of optical character recognition (OCR) was necessary to capture the text from these documents, so that text could be searchable for discovery purposes. Depending on a variety of factors, such as image clarity and how skewed it was when the image was taken, the quality of the OCR output varied widely.

As the use of email platforms became standard for communication within the workplace and work product was being created using Microsoft Office tools like Word, Excel and PowerPoint, it became evident that it was far more useful to work with this evidence in its digital form in discovery, which greatly reduced the need to scan and OCR documents. Early eDiscovery solutions were designed to mimic paper-based review, rendering emails, Word documents and other sources of ESI into static TIFF or PDF files. With most ESI formats resembling documents, it's not surprising that the legal community continued to conform evidence into a document-based structure.

However, as more types of data entered the picture – such as text messages and other data from mobile devices, chat messages, audio and video files, and databases – and/or became common sources of ESI in discovery. These formats much more resemble data, not documents, yet the legal community has "doubled down" on forcing them into the document paradigm. Modern eDiscovery platforms, despite their growing sophistication over the years, have continued to perpetuate this paradigm. Nearly every piece of digital evidence—whether a text or chat thread, a cloud-based collaboration exchange, or a collection of data from a database—is converted into a "document" for linear review.

For example, text and chat conversations are typically converted into 24-hour "chunks" (which often only reflect part of the conversation) or the entire chat thread (which often include other conversations that are not relevant to the case). ESI from databases may be limited to information available in pre-canned reports which could overlook important facts. This artificial structuring strips away the interactive and contextual elements of digital data, forcing legal teams to analyze evidence in ways that do not align with how it was originally created, stored, or used. The result is an incomplete and often misleading representation of the facts.

Other industries have fully embraced the digital age by leveraging data analytics to drive efficiency, innovation, and decision-making. For example, healthcare has integrated AI-powered diagnostic tools and predictive analytics to improve patient outcomes, streamline operations, and personalize treatment plans. The financial sector has adopted advanced data analytics and machine learning to detect fraud, assess risk, and deliver hyper-personalized banking experiences. **The healthcare and financial industries were both historically document centric, but they have changed the paradigm to data, not documents—recognizing data as a strategic asset, using technology to transform workflows, enhance customer experiences, and gain a competitive edge in an increasingly digital world.** It's time for the legal industry to do the same.

Drivers for Changing the Document Paradigm

The drivers to treat evidence as data, not documents have been growing for almost as long as eDiscovery has been a discipline! Here are three of those drivers.

Data, Not Documents

Unlocking the True Value of Your Digital Evidence

2006 Amendments to the FRCP

The 2006 amendments to the Federal Rules of Civil Procedure (FRCP) were a major step toward thinking of evidence as data, not documents. Among the changes, they significantly impacted eDiscovery by clarifying obligations related to the preservation and production of ESI. Rule 34(b) specified that parties must produce ESI as kept in the usual course of business or in a "reasonably usable" form, preventing the production of data in formats that hinder review and analysis. Additionally, Rule 37(e), which was updated in the 2015 amendments to the FRCP, introduced a framework addressing sanctions for failing to preserve ESI when litigation is reasonably foreseeable, emphasizing the importance of defensible preservation practices. These amendments reinforced the necessity for organizations to implement structured data management and preservation policies to ensure compliance and mitigate legal risks.

Volume and Variety of Data

The explosive growth of data in the world has been another major driver to treat evidence as data, not documents. According to Statista, data in the world has grown from **2 zettabytes** (2 billion terabytes) in 2010 to **182 zettabytes** by 2025 and is expected to continue to grow to **394 zettabytes** by 2028! With so much data in the world, it's become paramount in eDiscovery workflows to efficiently treat data in its native form and not convert it to static image forms which drive up data volumes even more—in turn, driving up hosting costs considerably.

It's not just the volume of data that's at issue—it's also the increasing data velocity and variety that has led to the Big Data world we all live in today. As demonstrated in the Internet Minute infographic from eDiscovery Today and LTMG, the ever-increasing variety of data types that bear no resemblance to documents—which includes everything from text messages to Microsoft Teams chats to audio and video files to emojis to prompts and responses on ChatGPT—illustrates the growing need to treat evidence as data, not documents. It's only going to get more diverse over time.

Explosion of Generative AI and Analytics

Speaking of ChatGPT, the explosive growth of the use of generative AI and analytics tools have been another driver to treat evidence as data, not documents. The tools available today enable organizations to obtain insights into their data like never before. The "dumbing-down" of granular data into a static, inefficient document format like TIFF or PDF destroys the ability for organizations to maximize the value of that data—it's like having a Maserati with no gas to put in the car!

Understanding Evidence Today

To maximize the value of evidence today, it's important to understand that evidence and how to work with it in the most efficient way possible. Digital evidence is either structured or unstructured; however, several forms of unstructured data are stored in a structured database container which enables that data to be analyzed at a granular level. In this section, we'll define and differentiate structured and unstructured data, discuss considerations in working with structured data, and discuss promising legal industry initiatives to drive the industry to treat evidence as data, not documents.

Definitions

To understand the difference between structured and unstructured data, some definitions[1] are in order:

Structured data is highly organized and formatted in a way that makes it easily searchable within databases. It typically resides in relational databases and follows a predefined schema, such as tables with rows and columns. Examples of structured data include financial transactions in a banking database, customer records in a CRM system, and inventory management data in an ERP system.

Unstructured data, on the other hand, lacks a predefined format and does not fit neatly into traditional databases. It consists of diverse content types such as text, images, videos, and emails, making it more challenging to organize and analyze. Examples of unstructured data include emails, social media posts, text and chat messages, and multimedia files like videos and voice recordings.

The Big Myth Regarding Unstructured Data

If you've ever worked with data in a Microsoft Excel workbook or Google Sheets, you know that there is an inherent organization to that data. Data is organized in rows and columns—making it easy to organize that data efficiently. Multiple sets of organized data with overlapping characteristics can be stored within multiple sheets within an Excel workbook; for example, a spreadsheet of invoices could pull data from other sheets in the workbook containing data about the customers buying products and the product inventory.

Great for Excel, but that doesn't help when working with data types like text messages, right? Wrong. **The big myth about unstructured data is that there is no structure to it at all, while the reality is that many unstructured data types are contained within structured databases and other container files.**

For example, let's look at text messages. While text messages are considered unstructured data due to their free-form nature, they are typically stored within a structured SQLite database on mobile devices. SQLite is a lightweight, self-contained relational database management system that organizes data into tables with rows and columns.

[1] We're not going to lie – we got these from ChatGPT. 😊

Data, Not Documents

Unlocking the True Value of Your Digital Evidence

For example, on both Android and iOS devices, SMS and messaging app data (such as iMessage or WhatsApp chats) are stored in SQLite databases. These databases contain structured fields such as:

- **Message ID** (unique identifier for each message)
- **Sender and recipient details** (phone number or contact ID)
- **Timestamp** (date and time the message was sent or received)
- **Message body** (the actual text, which remains unstructured)
- **Status flags** (e.g., sent, delivered, read)

"Text messages are stored in a database?!? Who knew?"

While the storage structure is highly organized, the content of the messages themselves remains unstructured. This combination allows for efficient searching, retrieval, and indexing while still dealing with free-form text data. Advanced analytics tools can extract insights from these databases by applying natural language processing (NLP) techniques to unstructured message content.

Text messages aren't the only example of unstructured data that's stored within structured container files—there are several other examples as well:

- **Emails in a PST or MBOX File** – Email messages, which contain unstructured text, attachments, and metadata (e.g., sender, recipient, timestamps), are stored in structured container formats such as Microsoft Outlook's PST files or MBOX files used by various email clients.
- **Documents in a Document Management System (DMS)** – Systems like SharePoint store Word documents, PDFs, and presentations inside structured databases, where metadata (such as author, creation date, and document type) is structured, while the actual content remains unstructured.
- **Multimedia Files in Databases** – Images, videos, and audio recordings are often stored as binary large objects (BLOBs) in relational databases (e.g., MySQL, PostgreSQL). Metadata (such as file size, format, timestamp, and geolocation) is structured, but the content remains unstructured.
- **Transcribed Voicemails in CRM Systems** – Customer Relationship Management (CRM) platforms like Salesforce and Zoho store voicemail recordings in structured databases, with metadata (caller ID, timestamp, duration) organized, while the audio content remains unstructured.
- **Chat and Messaging Data in SQLite** – Text messages aren't the only unstructured data format stored in SQLite. Applications like WhatsApp, Microsoft Teams, and Slack store chat conversations in SQLite databases as well. While the database is structured with tables for sender, recipient, and timestamps, the chat content itself remains unstructured.

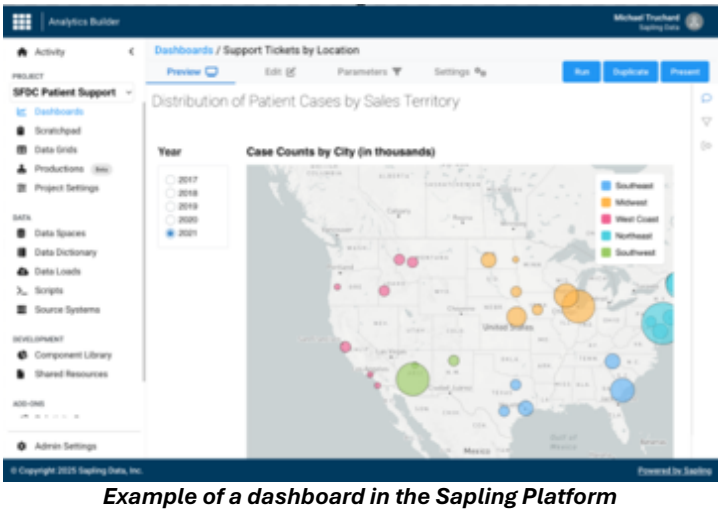
These examples illustrate how unstructured data is contained within structured environments, enabling better management, searchability, and integration with analytical tools. It's data, not documents, "kept in the usual course of business" as discussed in Rule 34(b) of the FRCP.

Benefits of Working with Structured Data

Are there benefits of working with structured data, not documents in discovery? Yes! Here are some of those benefits:

- **Consistency and Accuracy** – Data stored in structured formats follows strict data integrity rules, ensuring reliable results when querying.
- **Automation and AI Readiness** – Machine learning, predictive coding, and AI-driven analytics perform better on structured datasets because the data is typically clean and well-organized.
- **Faster Processing Speeds** – Queries on structured data run much faster than full-text searches in unstructured documents, reducing processing time and costs.
- **Regulatory Compliance and Audits** – Structured data is easier to track, log, and audit, which facilitates compliance with regulations like GDPR, CCPA, and SEC rules.
- **Data Linking and Correlation** – Structured data allows legal teams to connect related information across different data sources (e.g., linking email metadata with HR databases to track employee departures before key document deletions).

Structured data significantly also enhances data queries, visualization, and dashboarding, making early data assessment (EDA) in eDiscovery more efficient. Because structured data is stored in organized relational databases, legal teams can quickly run targeted searches using SQL or other query languages. This allows them to filter emails by date, sender, or recipient, search financial records for specific transaction amounts, or analyze metadata fields such as document creation timestamps and access history. Compared to unstructured data, which requires extensive manual review or advanced text-based searching, structured data enables fast and accurate retrieval of relevant information.



Example of a dashboard in the Sapling Platform

Data visualization and dashboards further enhance early assessment by providing interactive, real-time insights into key metrics. Structured data can be transformed into timelines of communications, heatmaps of data volumes, and pie charts categorizing document types, all of which help legal teams quickly identify trends, anomalies, and potential risks. Dashboards powered by structured data allow for real-time tracking of custodian data volumes, data source distributions, and privilege keyword hits, ensuring that teams can prioritize review efforts and refine search terms efficiently.

By leveraging structured data in this way, organizations can streamline their eDiscovery processes, reduce costs, and gain valuable insights early in the litigation lifecycle.

Overcoming Objections to Working with Structured Data

We get it—people are generally resistant to change, especially when that change could impact how they respond to discovery requests, how they bill for their services and whether there is sufficient return on investment to justify the change. Here are some of the objections to working with structured data and how to overcome those objections using a platform like Sapling.

"I don't want to produce structured data to the other side"

Objections regarding production of structured data have been around as long as eDiscovery has been a discipline. Remember the example of converting Excel files into a static image format like TIFF files for production. Legal professionals learned that it wasn't a useful way to produce them, and it didn't conform to Rule 34(b) as it was neither in the form in which it was maintained or in a reasonably usable form. Today, it's standard practice to produce Excel files in their native form with a single Bates number to represent the entire file. The same approach could be applied to any other structured data collection.

Having said that, you're not required to produce in the structured data format, and you can still obtain the benefits of analysis of structured data to cull and reduce data volumes, then convert remaining potentially relevant data (e.g., targeted text message conversations) to a document format for production.

"I don't know how to work with structured data"

The analytic tools available today make working with structured data easier than ever before. The natural language capabilities of generative AI enable inexperienced users to find the information they need without a lot of training. In cases where more expertise is needed, you may already have a database expert or Excel/Google Sheets power user experienced in working with structured data (most organizations do). If not, there are plenty of experts out there who can (including the Sapling Data team). You already use experts for things like forensic collection of mobile devices and predictive coding—this is simply one more discipline that can benefit from expert assistance.

"If I use an expert, I won't have visibility into the data analysis"

One of the benefits of using a platform like Sapling is that you and your expert can share access and analyze the data at the same time. The ability to direct an expert to search for specific data that relates to your case and watch your expert do it in real time enables you to collaborate in a way to quickly ensure you understand the data you have and how it impacts your case.

"Working with structured data is too expensive"

Hosting structured data and retaining an expert to help analyze that data is an expense. But when you consider the costs associated with processing entire structured data collections into static documents, as well as hosting and reviewing those documents, the cost savings associated with early data

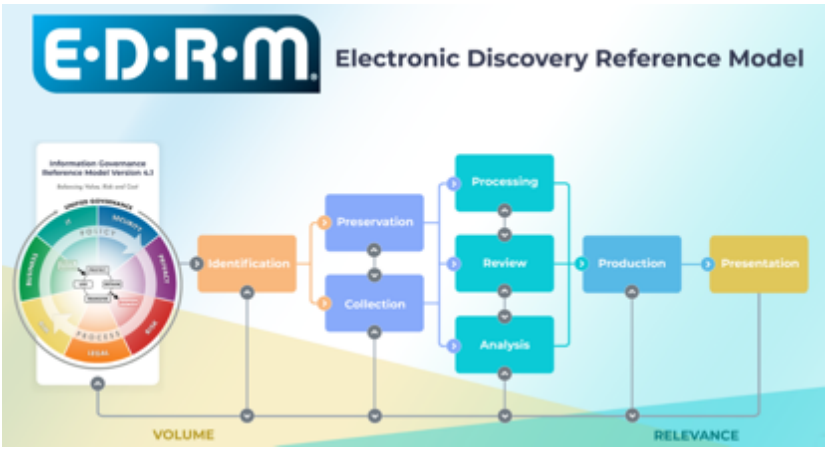
assessment of structured data and performing a targeted collection from that data (reducing those processing, hosting and review costs) typically outweighs the expense of working with the structured data, providing a tangible return on investment. And the additional insights obtained in analyzing the structured data may significantly impact your case, putting your team in a better position to make strategic decisions as the case unfolds!

Data Oriented Industry Initiatives

While we have discussed how the legal industry is stuck in the "document paradigm", there are promising initiatives toward a future where the industry is focused on data, not documents. Here are two examples.

EDRM 2.0 Project

Since 2005, the iconic Electronic Discovery Reference Model (EDRM) has been a standard illustration of the phases involved in a typical eDiscovery life cycle. While the model is a terrific illustration of eDiscovery activities, and is especially helpful to newcomers in the industry, changes in how eDiscovery is being conducted have caused EDRM to launch an "EDRM 2.0" project to update it to reflect emerging use cases, new data types and new technologies.



Current EDRM Model

This is a prime opportunity for the legal industry to break free of the document paradigm and consider approaches that emphasize data, not documents. Sapling Data is a proud participant in this initiative and a new EDRM 2.0 model is expected to be published soon for public comment.

Legal Data Intelligence Initiative

Launched in 2024, the Legal Data Intelligence (LDI) initiative is designed to empower the legal industry with a vocabulary, framework, and best practices to manage legal data. LDI covers the processes and best practices for use cases in litigation and investigations, regulatory requests, data breach response, and more – directly impacting. The processes represented by the Legal Data Intelligence model are designed to weed through the Redundant, Obsolete, and Trivial (ROT) data to enable the Sensitive, Useful, and Necessary (SUN) data to shine through, illuminating the information so it can be confidently transformed by legal professionals into new levels of intelligence. This focus on legal data to streamline many of the use cases and workflows associated with eDiscovery shows promise for getting the industry to focus on data, not documents.

What Working with Structured Data Looks Like

What does working with structured data look like? As you'll see below, it involves the execution of key eDiscovery tasks tailored to the nature of structured data, applied to real-world use cases where discovery of structured data enables legal teams to understand their evidence at a much more granular level because that evidence is in the form of data, not documents.

eDiscovery Tasks for Structured Data Workflows

Like any other data source, the lifecycle for discovery of structured data involves eDiscovery tasks and phases you're familiar with, but the execution of those tasks and phases is tailored to the nature of the data.

Collection

Data collections are most commonly formatted as CSV (comma-separated values) text files. Although a CSV file may appear similar to a standard document, its structured nature sets it apart. Unlike typical documents, CSV files can contain millions of records, each organized into distinct fields. The first line of a CSV file, known as the header, defines these fields by assigning names such as first_name, last_name, city, state, and zip. Each subsequent line represents an individual record, with values separated by commas—for example, "John", "Smith", "Austin", "TX", "78757". If a file named "patients" contains 10,001 lines, it holds 10,000 individual patient records (excluding the header).

Data collections may also originate from database backups, system-generated reports, or API extractions—though APIs are used less frequently due to security concerns. Regardless of the method, the result is a structured dataset consisting of tables with rows of data, making it easier to process, analyze, and integrate into eDiscovery workflows.

De-Duplication

Document collections aren't the only data sources that can contain duplicates – they are also common in structured data sources as well. It's easy to identify and eliminate duplicates in a structured data file once it is loaded into a database for analysis.

Culling & Filtering

A key advantage of working with structured data is the ability to filter and refine datasets using specific fields, allowing for efficient data review and analysis. For example, if a dataset contains a client's order history, but only orders from 2022 are relevant, a field like "order_date" can be used to exclude records outside the specified range, focusing only on responsive data. Similarly, in an investigation involving 500,000 insurance claims, where only three specific procedure codes are of interest, a designated procedure field can be used to isolate relevant claims while filtering out unnecessary records. This functionality is possible because structured data is directly tied to a client's operational records, which are systematically gathered and maintained within business systems. Preserving these fields during data

collection allows for precise filtering, improved insight, and greater efficiency, ultimately driving innovation and productivity in eDiscovery workflows.

Data Visualization and Analysis

Example of Data Visualization in the Sapling Platform

Data visualizations help illustrate quantities, relationships, proportions, timelines, and other key dimensions of information, providing context and highlighting the significance of the data. The structured nature of data, with its discrete fields, enables the creation of a wide range of visual representations. Charts, maps, timelines, and diagrams serve as foundational elements, and when enhanced with filters, color coding, comparisons, and statistical metrics, they transform raw data into compelling insights. By leveraging these tools, legal teams can turn complex datasets into a strategic narrative, making it easier to uncover patterns, assess risks, and support legal arguments.

Search

Searching structured data is both powerful and highly flexible due to its organized format with clearly defined fields. Users can choose to search across all fields, specific fields, or a single field, depending on their needs. Additionally, databases support pattern recognition, partial term searches, and complex filtering conditions, allowing for precise queries. For example, a user could retrieve all insurance claims referred to Dr. Jones by Dr. Smith between January 1, 2020, and September 15, 2021 by applying multiple search criteria simultaneously. These capabilities make it easier to pinpoint relevant information while efficiently filtering out non-responsive data, ensuring a more effective and targeted eDiscovery process.

Review

Reviewing structured datasets—even large datasets—is often accomplished by dividing the data into subsets or cohorts (i.e., groups of data) based on shared characteristics. For example, a dataset might be filtered to identify all mobile contractors in Oklahoma who visited more than five patients per day but filed fewer than four travel logs per day in their timekeeping reports. Defining such a group using dataset fields as parameters makes the cohort both logically explainable and quantifiable, allowing for an exact count of individuals meeting the criteria. From there, legal teams can easily analyze who is involved, when the activity occurred, and how many records fit the pattern, which is essential for assessing risk, estimating damages, or identifying outliers. The query capabilities of structured data supports rapid hypothesis testing and interactive dashboards, enabling legal teams to dynamically explore and refine their findings in real time.

Redaction

Because structured data is highly organized, redactions of sensitive data—including personally identifiable information (PII)—is much more easily performed, as that data is often isolated to specific data fields, like name, address, social security number (SSN) and more. Replacing that data with patterns (e.g., ***-**-**** for SSN) is easily performed in a single command.

Production

As discussed above, analyzing a structure data set doesn't mean it's required to be produced; however, when it's appropriate to produce a structured data set, the file (or files) being produced should receive a single Bates number (just as you're currently doing with Excel files that are produced). Produced structured data sets should also contain a unique identifier for each record—which is typically called a "Primary Key"—that enables the Court and counsel to refer to specific records within the set, just as Bates number does with specific documents.

Use Cases for Working with Structured Data

There is a virtually unlimited set of applicable use cases for working with structured data. Hopefully, the examples below illustrate the importance of working with this evidence as data, not documents.

Sales and Customer Support Data

Analyzing structured data from customer support tickets provides valuable insights into how a client manages customer interactions, resolves issues, and maintains service quality. By organizing support data into structured fields—such as ticket ID, issue type, customer ID, time to resolution, assigned agent, and customer satisfaction scores—legal and business teams can systematically assess response efficiency, resolution times, and recurring complaint patterns. Structured data allows for trend analysis, helping to identify bottlenecks, agent performance disparities, or gaps in service protocols that may impact customer retention and compliance with service-level agreements (SLAs).

Additionally, structured datasets make it easier to correlate support performance with sales trends, uncovering whether poor customer support experiences contribute to lost revenue or increased churn. In legal contexts, structured customer support data can be essential for investigating claims of negligence, breach of contract, or regulatory violations, ensuring that organizations maintain transparent, measurable, and defensible service practices. By leveraging structured data, businesses can refine customer support workflows, enhance operational efficiency, and proactively mitigate risks.

Medical Billing Data

From a healthcare perspective, medical billing data can be a key data source to identify potential billing mistakes or even potential instances of fraud. For example, analyzing a structured data set of medical billing data can identify instances where the wrong amount was billed because the incorrect geographical location was applied—such as billing through the wrong Medicare Administrative Contractor (MAC) based on its geographical location. Incorrectly sending lab monitoring data to an offshore lab is another example

of the type of issues that can be identified within a medical billing dataset. Overpayment to a specific MAC could be a sign of fraudulent activity and could result in damages and penalties under the False Claims Act (FCA).

Wearable Medical Device Data

Analyzing structured data from wearable heart monitors is crucial for assessing whether monitoring centers are responding to cardiac events in a timely and effective manner. These devices continuously collect heart rate, rhythm patterns, and other physiological metrics, generating timestamped data logs that can be analyzed to detect abnormalities such as arrhythmias or sudden drops in heart rate. By working with structured data from these devices, legal and medical teams can track response times, measure delays between event detection and intervention, and identify patterns of delayed responses or missed alerts. For example, a structured dataset can help determine whether patients experiencing critical heart events received timely alerts and medical intervention based on predefined escalation protocols. Additionally, structured data allows for benchmarking against industry standards, ensuring compliance with regulatory requirements and identifying potential liabilities in cases of negligence or malpractice. By leveraging structured data for analysis, organizations can improve patient outcomes, mitigate legal risks, and enhance the overall effectiveness of remote monitoring systems.

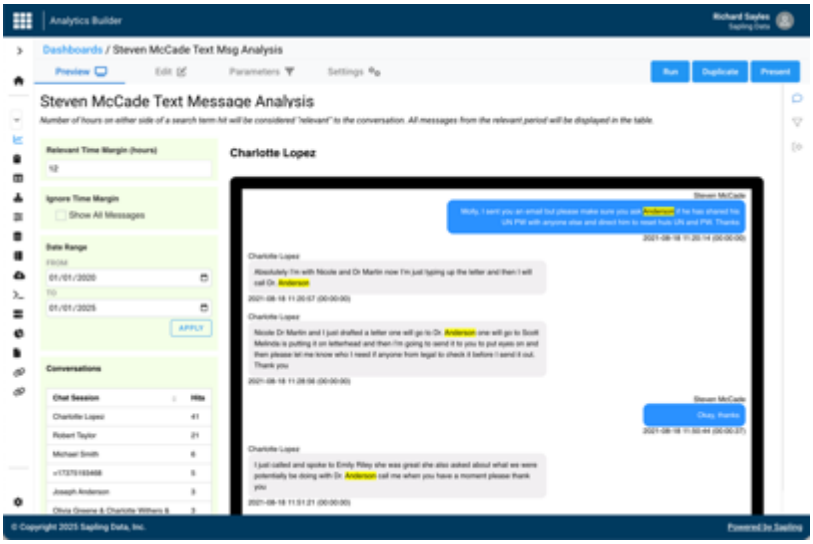
Example of Medical Device Data in a Grid View

Data on Employment Disputes

Structured data can be highly important in an employment dispute by providing clear, quantifiable evidence related to workplace activities, policies, and compliance. Data from HR systems, payroll records, timekeeping logs, performance evaluations, and email communications can be analyzed to assess claims involving wrongful termination, discrimination, wage and hour violations, or workplace harassment. For example, timecard data can help verify overtime claims or missed break allegations, while structured email metadata can establish patterns of communication that support or refute allegations of a hostile work environment.

Additionally, structured performance review scores and promotion records can help determine whether hiring or termination decisions were based on merit or if they reveal potential biases. By leveraging structured data, legal teams can identify inconsistencies, measure trends, and present objective, data-driven insights, ultimately strengthening the case for either the employer or the employee.

Text Messages



Example of Text Message Review in the Sapling Platform

While the examples above are related to specific types of claims and cases, the discovery of text messages has become common across all types of cases. The arbitrary way in which many eDiscovery providers convert those text messages into PDF files in 24-hour "chunks" eliminates the flexibility of working with the text messages individually to not only pinpoint specific relevant and responsive conversations but also the specific relevant and responsive messages within those conversations.

A common workflow is to extract the text message corpus from a specific custodian Into a CSV file, then load that into a platform (such as the Sapling platform) for analysis and review. Target queries within the platform can quickly identify potentially responsive individual messages, which can be reviewed, along with other messages to tag the relevant and responsive conversations containing those messages. Those tagged messages can then be exported and loaded into an eDiscovery review platform to be produced, along with other relevant ESI. This ensures that only relevant and responsive messages are produced within relevant and responsive conversations, eliminating irrelevant data productions and minimizing privacy concerns.

Conclusion

Do you still rent movies from Blockbuster Video? Of course not—streaming technology has advanced to the point that physically going to a store and renting a movie is obsolete. The "dumbing down" of structured data by forcing it into the document paradigm that legal teams have been accustomed to—for centuries, literally. Working with structured data, not documents in discovery allows for precise filtering, rapid analysis, and interactive insights, eliminating the inefficiencies of forcing it into a document-based review process that obscures relationships, patterns, and quantifiable trends. **Working with data, not documents not only leads to reduced discovery costs but also promotes strategic planning for the case overall.**

So, when do you need to consult a structured data expert and/or consider using a dedicated platform for analyzing, searching and reviewing structured data? Anytime you have structured data for which targeted analysis of that data could benefit your case—including for text messages, which have essentially become ubiquitous in eDiscovery collections. The benefits of working with structured data, not documents, in discovery are clear—now, it's up to you to explore those benefits. It never hurts to ask!