

LEGAL TECH

Building Blocks for Efficiently Handling Structured Data in E-Discovery

By David Turner and Colleen Casey Voshell

To attorneys, the mention of structured data could bring forth feelings of indifference, confusion, anxiety, frustration and other mixed emotions. For some, the task of simply understanding the nuances between structured and unstructured data can be a daunting endeavor. That said, there is a wealth of information about many data sets living as structured data (i.e., data residing in databases—think tables and fields, not documents) within a variety of systems. When tackled strategically and with the right approach, leveraging structured data can increase e-discovery efficiency and provide valuable information about document sets that might otherwise be invisible or lost.

As an example, consider a product liability suit that involves collection, analysis and review from massive data repositories, including product quality assessments, compliance reports, regulatory reporting systems, and other documents that reside in a variety of databases or systems within an organization. In these instances, counsel is faced with a data set of hundreds of thousands of documents—hundreds of gigabytes of data—that need to be reviewed. Traditionally, counsel approaches these data sets by using keyword searching to locate and collect the potentially relevant data for review. However, by leveraging structured data from those systems as the lens through which to look at the documents and data, attorneys can identify if the relevant product issue or defect has a unique identifier (e.g., a certain code, ID number or other classifier that may even be constant across the enterprise). If so, rather than using keywords that might be under and/or over-inclusive, the team can query the structured data identifier to find all of the records and

underlying documents associated with that product or issue.

Counsel can further refine this methodology to find specific date ranges, metadata of relevance, document title or category hits or other clues to unlock potentially relevant information from within the data set. Counsel can use this process to exclude documents from the data set as well, and potentially reduce the overall review set—which results in saved time and money for attorneys and their clients. This type of analysis can be particularly useful in industries that rely heavily on enterprise-wide system integration, including financial services, insurance, health-care and manufacturing.

Below we outline a few key considerations for successfully incorporating structured data into the early stages of an e-discovery matter, and discuss how to ensure efficiency and proper handling of structured data. Keeping these considerations in mind, attorneys can determine which ones apply to an individual case, and develop strategies accordingly.

1. Understand the Systems and Data Types

It is impossible to effectively use structured data for e-discovery without first understanding what types of systems house it, where these systems exist within the enterprise, what type of information is there and who the key business users are. Any systems that include both documents and structured data components, such as those used for invoicing, contract management, schematics and engineering/design, are often sources of potentially relevant information. Documents stored in a document management system are often saved with user-created profiles that provide information about the documents; claims files may be tied to systems that track customer calls. As an example, we can look at activity within a financial institution, wherein Bloomberg chat data can be parsed by using customized structured data tools so that the chats are distinct records rather than long strings of text. Counsel can review and analyze these parsed chats to identify chats that might signal questionable conduct. Then, rather than taking a broad and undisciplined approach to looking at a wide range of trade activity, attorneys can focus only on trades during the time period or relating to the activities identified through the analysis of the Bloomberg chat data to locate specific activities in question.

Gaining understanding—essentially creating the system data map—involves both business and technical discussions with stakeholders in IT, legal and other departments that use these systems, to understand their business purpose and the scope of the data that may need to be collected from them.

Internal stakeholders should work with counsel and technical consultants to help shape the process and guide what can be done with the information. This will allow the e-discovery team to cast a narrow net and sharpen the focus of the collection and subsequent review.

2. Discern How the Systems are Used

Each organization and business department is unique in how it is using its systems, what reports from the system they rely upon, and how information can be searched. These individualized processes often do not exist in prepackaged user guides, but significantly impact how the information is stored and can be searched. While knowing what is in the system is an important first step, the requirements of a data collection will be defined by user practices, and an attorney cannot execute a collection without a meaningful understanding of those practices.

Obtaining this level of detail is best accomplished by talking directly to business users and, as mentioned above, relying on them as internal subject matter experts. Internal IT and/or technical consultants should be involved as well, to ensure that the technical capabilities of the system are covered in these discussions and any technical limitation, or solution, is properly vetted and discussed. A practical example is an attorney working on an insurance matter who is specifically looking for claims associated with a particular code. By understanding how the insurance processor uses that specific code, and the technicalities of how to identify the relationship between it and where the associated documents are stored, the attorney can access and search the documents. This would be a fool's errand for anyone without a sound understanding of the system and how it is used.

3. Know the Right Time to Leverage Structured Data

With some systems, analyzing structured data to determine relevance first can bring significant efficiencies to the review because then counsel might only need to collect and review the documents identified as potentially relevant through the structured data analysis. In other systems, it makes more sense to look at the documents and structured data in tandem, as sometimes neither can stand alone in identifying relevance. In still other systems, the structured data may not play a role in relevance, or you can satisfy your discovery obligations simply by producing a report generated from the structured data system. Identifying the right point in

the process to involve the structured data will help inform the strategy and methodology for handling these data and document sets and will ultimately reduce the amount of data collected, reviewed and produced, and lead to better control of total e-discovery costs.

These decisions are key in establishing the review design and will impact how an attorney funnels data through the e-discovery process. For example, when querying a document management system containing five terabytes of data, the structured data volume is going to be significantly smaller than the document volume. In this scenario, the attorney can start by systematically querying and analyzing the structured data first, thereby creating an e-discovery process that involves only collecting, reviewing and producing a small subset of that mostly document-based, five terabyte system. In a second scenario, the attorney might instead take a targeted extract of the structured data, have that reviewed by contract attorneys, and determine relevance that will guide which additional data must be collected and reviewed for responsiveness.

In any matter, it is important to keep in mind that reviewing structured data is a technical process that requires specialized tools and expertise and an analytic approach. It might open up a level of disclosure with opposing counsel beyond some attorneys' comfort zone. You can avoid many e-discovery headaches by preparing yourself, before initial disclosures and/or a meet and confer, with the technical knowledge of what systems may contain potentially relevant information, what data and/or documents are housed in those systems, how that information can be extracted and potential risks that could arise as a result. As with any intensive e-discovery project, the use of structured data in the collection and review phase should only be undertaken with the guidance and support of a technical expert that deeply understands the systems and has experience navigating these types of matters. •

For daily tech news, go to lawtechnologynews.com.

Turner is a senior managing director at FTI Consulting and is based in Washington D.C., where he leads the Financial and Enterprise Data Analytics group. Voshell is a managing director in the FTI Consulting Technology practice and is based in Philadelphia.