The Danger of Unstructured Data in Law Firms

By Sharon D. Nelson, Esq., John W. Simek and Michael C. Maschke
© 2022 Sensei Enterprises, Inc.

Unstructured Data – A Problem That's Been Around for a Long Time

Recently, authors Simek and Nelson had the opportunity to talk to guest Peter Baumann on their Legal Talk Network Digital Detectives podcast. Baumann is the CEO and founder of ActiveNav, a leading data privacy and governance software provider.

As far back as 2008, Baumann was observing the exponential growth of data and specifically unstructured data, the data that sits outside of databases. He noted that today, "the best technology, the best doors and locks and alarm systems won't stop the bad actors getting into your network. I think people understand that now." Data protection, data privacy, policies and regulations are crucial to employ so that, if your network is compromised, you are prepared by having the data correctly labeled, categorized, and locked down to reduce your threat attack surface.

The truth is you can't protect the data you don't know you have.

How is Unstructured Data Different from Structured Data?

If it's structured, it means that it's kind of already gone through some sort of filtering, triage and parsing system and it typically sits in some kind of managed structured environment such as a database. Think SQL, think Oracle.

Baumann told us, "If it's unstructured, I like to call it the wild west, it could be in potentially hundreds, maybe even thousands of different types of repositories from those that we're very familiar with, like our general office documents in the Microsoft or the Google stacks, through to a multitude of different tools that different organizations will use." Simply, what's not in the database is unstructured.

What are the most common examples of unstructured data?

Common unstructured data often consists of text, in many forms including text files and documents, Word documents, email messages (generally considered semi-structured data), text messages, PowerPoint presentations, survey responses, chats, transcripts of call center interactions and posts from blogs and social media.

Other forms of unstructured or semi-structured data types include images, video files, audio files, sensor data, server, website and application logs. Machine data is growing quickly – log files from websites, servers, networks and apps – particularly mobile ones. And have you thought to include data from IoT connected devices?

The percentage of unstructured data has been estimated at 80%. At a recent RSA conference, a renowned analyst said that they had done some recent work which found the percentage was even higher than that.

Just think about the numbers. In a world where data breaches are proliferating at an extraordinary rate, most organizations have 80% plus of their data in an unstructured environment and most of them have no idea what's in there.

Why is Unstructured Data a Big Threat to Law Firms?

As we've established previously, it's a risk to everyone but it may be more of a risk to law firms because of the nature of the data they are holding, including very sensitive and confidential information. Many times they also hold monies or access to monies.

The bad guys will sniff things out. When you've got an arbitrary collection of unstructured data sitting in an email account, or on a file server or chat stream for example, you've got no signals or tools to identify and manage that data. You're at risk. If hackers were to infiltrate the organization's network, potentially via an unstructured data source, there's nothing stopping them from getting hold of highly confidential client data files, court filings, contracts, deposition files, etc.

If a breach leads to a significant loss of confidentiality, that's huge. Trust and data protection is fundamental to the legal industry. Failure to keep data safe might be seen as unethical depending on the security measures taken. There are all sorts of compliance issues, with all the states and territories having data breach notification laws – and there are more and more privacy laws in place. The ultimate horror show is significant reputational damage, which could be devastating.

Why Do Law Firms Often Avoid Dealing with Their Unstructured Data?

Bauman said his flippant answer is that "it's just too hard." He added a slightly more nuanced answer, indicating they may not really understand the implications of what might happen and what's in the unstructured data. Dealing with unstructured data feels daunting. It can be a time-consuming and expensive process to deal with. Because most law firms are generally unaware of the magnitude of the risk, they procrastinate, perhaps not thinking of how expensive and time consuming a data breach would be.

They struggle with whether to do it now or push it off. They worry that they are not in control even of their structured assets. Some are only just realizing that perimeter protection is not sufficient and they are moving to a Zero Trust Architecture. But if firms ignore unstructured data, you may end up with thousands of potential entry points that attackers can exploit.

The return on investment for successfully dealing with your unstructured data is significant. In the early days, it was all about storage savings. Today, it's all about risk. Firms are always willing to deal with this post breach. But there may be something else that motivates them to act

sooner. Perhaps they are doing a data migration project and they are migrating data to a new cloud. That's often a good time to look at the data, clean it, label it, etc.

What Tools Should Law Firms be Using to Control Unstructured Data?

Baumann says too many people have had a bad experience with trying to control unstructured data. Maybe they used the wrong tool. Maybe they didn't even understand that there are tools to help. The key thing is you need tools that are built to do the job, not secondary or tertiary players in the market. You want to use tools that are built for unstructured data that have been built to have no knowledge of the content and build holistically ground up.

You don't want to pass the burden of making decisions about the data to the busy managing partners. Historically, a lot of organizations think that just carrying out a manual survey asking people about their data will suffice. It is an important part of the process, but it doesn't suffice. Once that survey is complete, it's out of date the next day. The other problem with surveys is your reliance on human recollection, which is ever and always faulty. So, you must combine those surveys with actual data.

For a listing of the top 15 Data Analysis Tools in 2022, check out this link: https://hevodata.com/learn/data-analysis-tools/.

How To You Get Buy-in From Law Firms to Put New Restrictive Policies in Place?

You show them their own data. Then you show them the risks they have within their firm. You run the out-of-the-box algorithms rules which will very quickly show information that shouldn't be there, personally identifiable information, non-compliant data, data that violates privacy law. You also show them the data against their existing in-house policies. For instance, Baumann recently talked to a law firm which never had a retention policy on their emails. So, they have an extremely high risk, having 20 years of emails.

If you have a policy that covers five years of emails, you've removed an enormous amount of risk by not having to deal with 20 years' worth of emails.

What Steps Can Law Firms Take Today to Govern Their Unstructured Data?

The first thing is you need a plan. Make sure you've got the right executive sponsorship, in other words, the managing partners. Then you get into the nuts and bolts. Do that survey of your people. That's a very helpful process to go through and most law firms, at least the larger ones, have probably already done it on their structured data, so leverage the same process across your unstructured data.

You need an up-to-date inventory of all your data assets. Once you have those in place, you can leverage the knowledge and experience and methodologies you've used already to your structured data and carry those across into the unstructured space.

You will need to bring the right tools in to support this process. That's going to vary depending on the size of the firm. You need tools to provide an up-to-date inventory of all your unstructured data assets. The inventory must be kept current. Align the inventory with your policies, procedures, and your other methodologies.

You essentially need to start an ongoing process of data remediation, management, and classification as appropriate for your law firm.

If your head is spinning, that is the appropriate response. Perhaps the first step is to get expert help and guidance!

Sharon D. Nelson is a practicing attorney and the president of Sensei Enterprises, Inc. She is a past president of the Virginia State Bar, the Fairfax Bar Association and the Fairfax Law Foundation. She is a co-author of 18 books published by the ABA. snelson@senseient.com

John W. Simek is vice president of Sensei Enterprises, Inc. He is a Certified Information Systems Security Professional (CISSP), Certified Ethical Hacker (CEH) and a nationally known expert in the area of digital forensics. He and Sharon provide legal technology, cybersecurity and digital forensics services from their Fairfax, Virginia firm. jsimek@senseient.com.

Michael C. Maschke is the CEO/Director of Cybersecurity of Sensei Enterprises, Inc. He is an EnCase Certified Examiner, a Certified Computer Examiner (CCE #744) a Certified Ethical Hacker and an AccessData Certified Examiner. He is also a Certified Information Systems Security Professional. mmaschke@senseient.com.