Your eDiscovery Data Repository is a Target.

Don’t worry. Here’s what to do about it.

An eBook for GCs, CISOs and their teams.

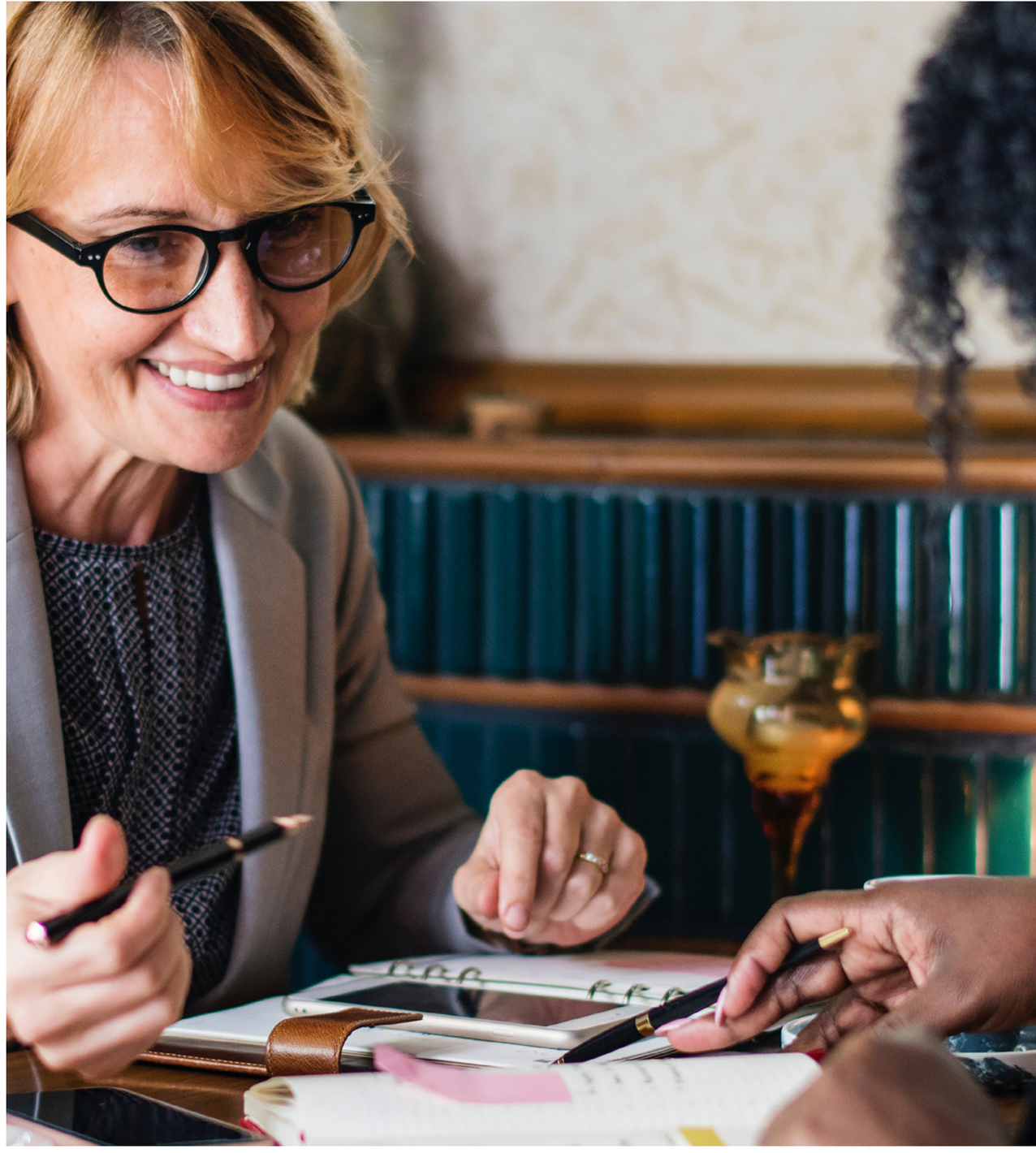
Table of Contents

Introduction

So, you're here to find out about your eDiscovery data repository. If you're a General Counsel, a CISO or just fascinated by the world of data security, we've got the skinny on why you might be at risk, and how to make sure you're protected in the future.

It's no secret that the number of corporate lawsuits is on the rise, and this means the amount of data you need to collect in response to eDiscovery requests is also growing. Consequently, there's a lot of gigabytes (or even terabytes!) of potentially responsive data to search for in just one lawsuit collection phase. Then, once you've found it, you've got to place it on litigation hold and review it before you produce it to your opposing counsel. That's a lot of info to have sitting around unprotected. You wouldn't let a stranger into your home, would you? So why let them have access to all of your sensitive data?

Very often, corporate legal departments resort to storing these massive datasets on a shared drive, without any additional security. This means you're relying on your existing enterprise security technology, which might not always be able to protect your data repository as well as it protects the rest of your systems.



So what's the real problem here?

The main issue is that eDiscovery datasets are an extremely rich source of sensitive company information. They can include trial evidence and attorney case preparation, compliance data, business strategies, M&A documentation, and even intellectual property. Based on our experience, these eDiscovery datasets and other sensitive legal information are almost never deleted, in case they need to be reused during appeals, regulatory information requests or other related matters, so they essentially remain in storage forever.

Over the last several years, business vulnerabilities in legal departments, law firms and service providers have led to hacking and theft, including theft of sensitive data by foreign governments. You're interested to know more, I'm sure; so are the Federal Bureau of Investigation (FBI) and other law enforcement agencies. The reason these vulnerabilities are so exposed is in no small part down to the lack of sophisticated security in so many enterprise storage systems, not to mention the constant evolution of technology and hacking techniques that cyber-criminals can now employ.

Alan Paller, the Founder and President of the SANS Technology Institute, shared an eye-opening conversation he had with the Managing Partner of a large New York law firm several years ago, about a hacking incident the law firm had had been alerted to by the FBI.

Alan Paller: What exactly did the FBI agents tell you?

Corporate Attorney: They said that our files had been found on a server in another country. The server was used as a way-station for sending data to a large Asian country. Off the record, they said it was China.

Alan Paller: Did they tell you which files?

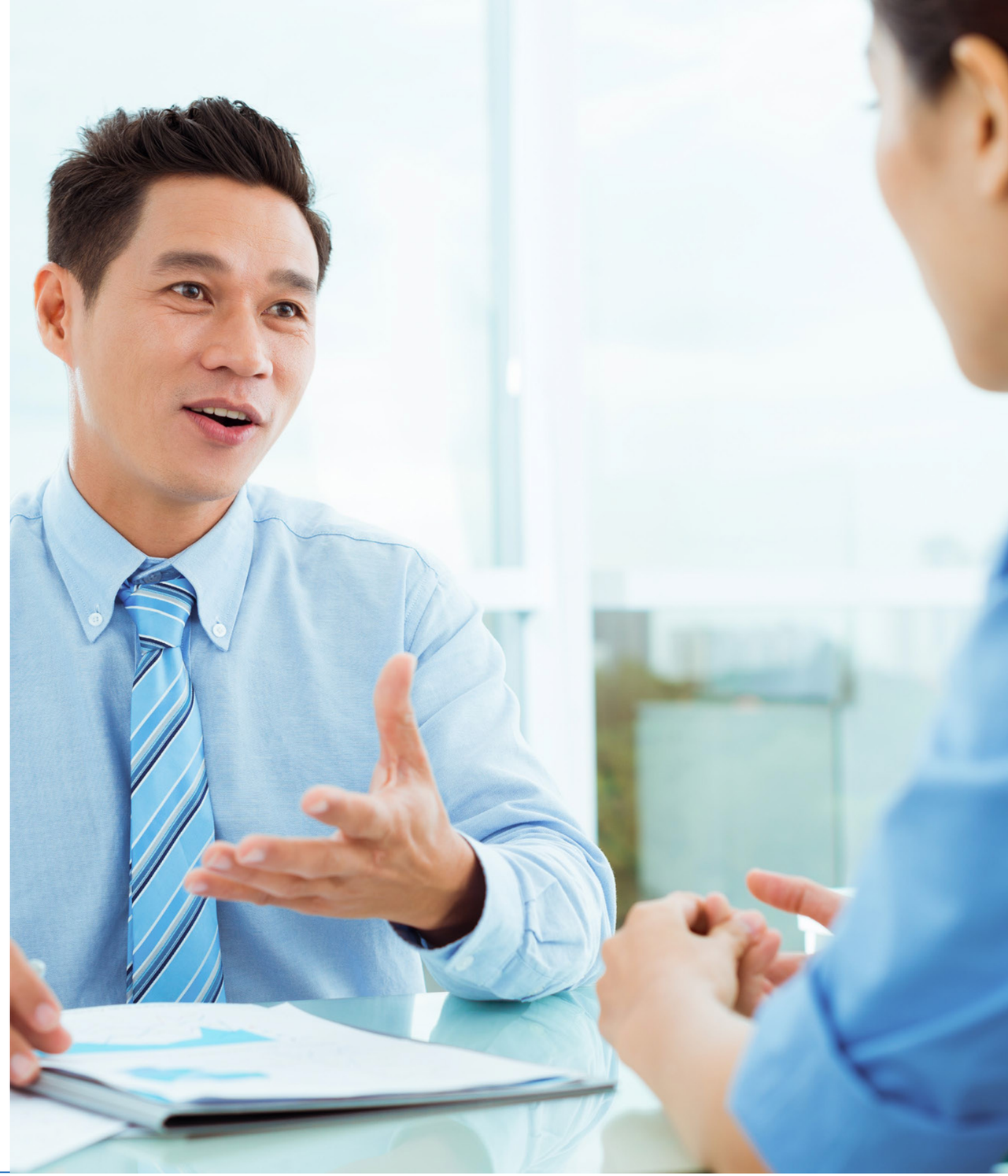
Corporate Attorney: They showed us a listing of what they had…It was all of our client files.

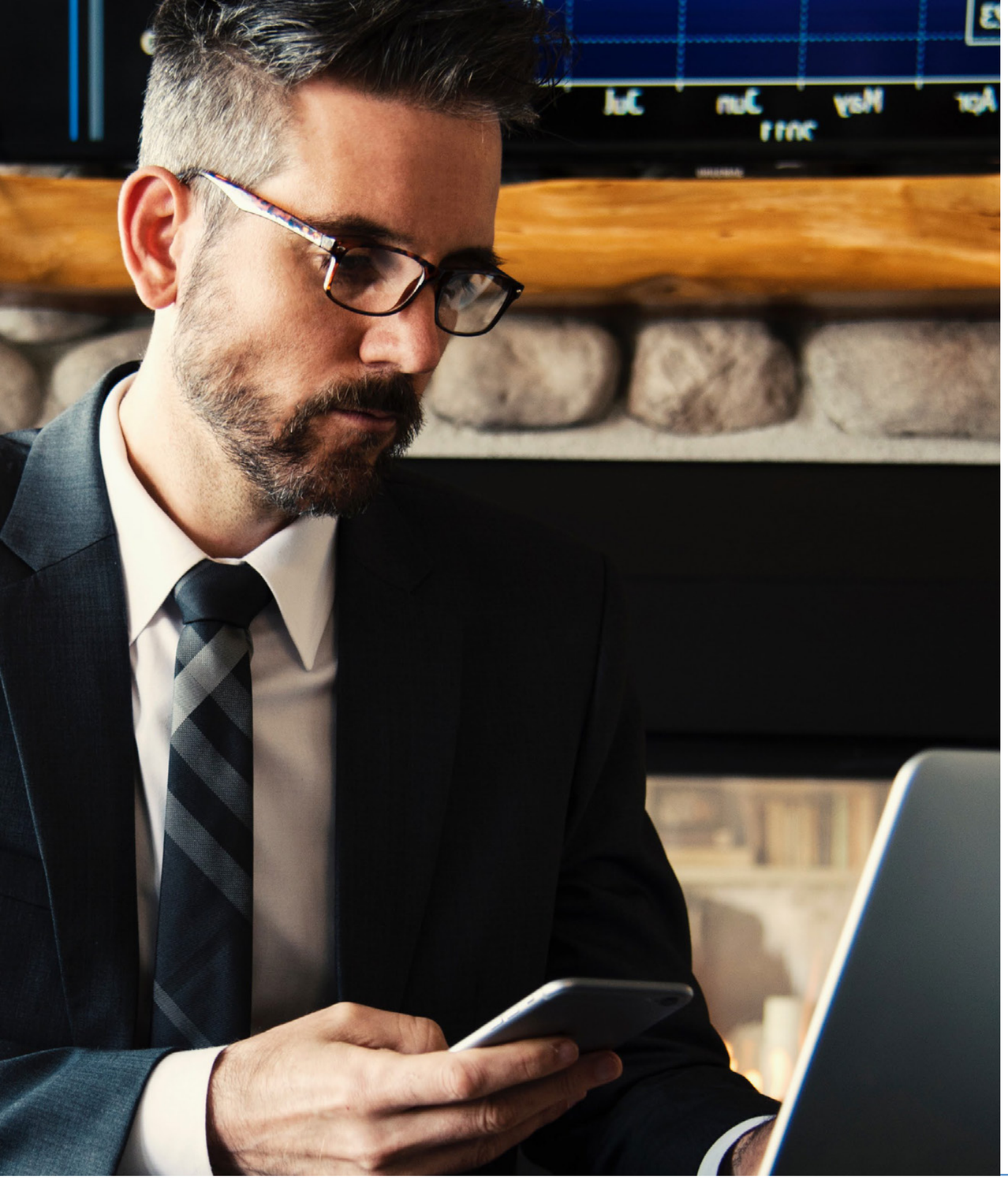
How does this apply to my organization?

We know that you and your teams care about security. We also know that deadlines are ever present and looming over you, which is probably why you continue to store huge amounts of eDiscovery data on file sharing systems. What you might not realize is that this security risk is made even worse when you provide your attorneys easy access and data portability, meaning they can work from their mobile phones and laptops when not in the office. Obviously everybody loves flexible working, but when sensitive data is stored on unsecure file sharing systems, there's a big risk to it. In fact, many industry pundits suggest that corporate legal department file sharing systems have experienced the same intrusions as law firms themselves.

So, you've got department file shares that are unsecured and potentially unmanaged, and you're relying on simple password access, with no audit or reporting capability. Chances are, most cyber-criminals would be able to gain access through a weak firewall and easily copy your entire file share. You probably wouldn't even notice. It's that frightening. What's worse is that these criminals can then review everything they've stolen and attempt to sell it to anyone who's interested, or even ransom it back to you. In many cases, you're unlikely to even know your datasets have been accessed and copied.

We're not denying the need for you to store your eDiscovery data for long periods of time, but if there was a more secure way to do it while keeping your budget as low as possible, wouldn't you want to know about it?





Worried? Don’t be, there’s an easy solution.

If you’re quaking in your boots, don’t start to panic just yet. These risks mean you’ll want to revisit your security capabilities when it comes to storing all this sensitive legal data.

There are two potential strategies you can choose from to address this:

1

Invest in upgrading and securing your existing system on-premises

2

Create new repositories in the cloud and migrate your existing

Unless you’ve been living under a rock (or hiding in your datacenter), you’ll probably know that the cloud is a pretty big deal. It’s not all just hype, either. There are a whole heap of security and business benefits to moving sensitive data to the cloud. In particular, cloud data lakes are emerging as a best practice for anyone looking to take advantage of the latest generation of technology and security.

Dip your toes in a data lake

What if your eDiscovery datasets were stored in a common repository that could be actively managed, searched, given the relevant legal holds, culled, reviewed and exported by you, your external law firms or even your opposing counsel whenever they needed to? Great news; they can be. The reality is that this common data repository, along with appropriate security and data management capabilities, already exists in the form of a cloud-based data lake.

Wondering what this means? In simpler terms, a data lake is a storage repository that can hold vast amounts of raw data in its native format until you need it. Each data element in a lake is assigned a unique identifier and tagged with a set of extended metadata tags. Then, when you need to, you can query your data lake to get all the relevant data. Once you've got the data you need, you can analyze the subset as required.

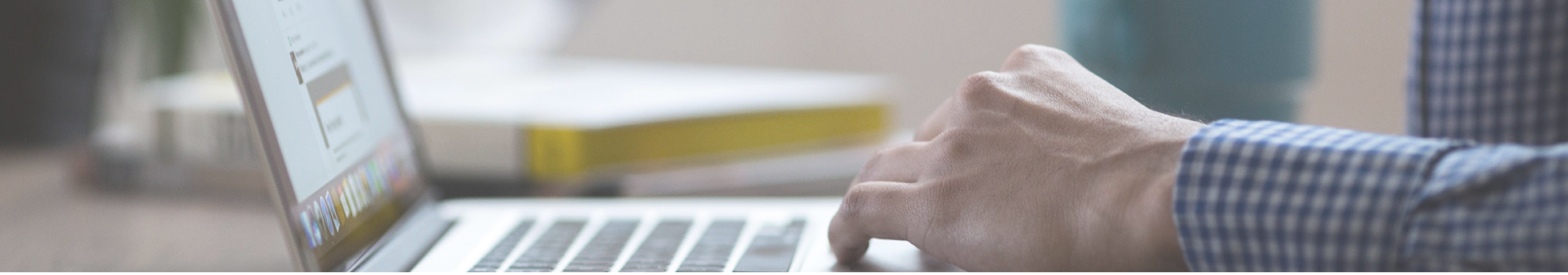
If you think of a data mart as a store of bottled water – cleansed and packaged and structured for easy consumption – the data lake is a large body of water in a more natural state. The contents of the data lake stream in from a source to fill the lake, and various users of the lake can come to examine, dive in, or take samples.

James Dixon, CTO Pentaho

Key attributes

Data lake technologies are offered by large, well known cloud infrastructure providers such as Microsoft Azure and Amazon Web Services, so the risk of vendor disappearance is greatly reduced.

- ➔ Data lakes adopt the cloud platform's constantly upgraded, next-gen security capabilities, including access controls, auditing and reporting.
- ➔ Additional functionality, like security and case management, can be added with the inclusion of additional software applications.
- ➔ A data lake is amorphous, meaning there's no set file format requirement or data management architecture that must be followed. Therefore, it's adaptable for any number of uses, including litigation support.
- ➔ All data and metadata is left in its native format.
- ➔ By default the data is unmanaged, but it can be actively controlled using additional software applications.
- ➔ A data lake has a single comprehensive search capability across all its data, speeding up search and production, and ensuring that the results are consistent.
- ➔ Data, both structured and unstructured, can be stored and indexed on the fly for more focused, complete search results.



Your employee has left the building…But their data shouldn’t.

All companies – yes, even yours – face the question of what to do with large datasets left behind by inactive or departed employees. When employees leave a company, they leave behind huge amounts of potentially valuable data, as well as data that’s subject to regulatory retention or eDiscovery requirements.

Data could be left on a shared drive, a company-supplied smartphone, an employee workstation or their Office 365 account (which includes OneDrive, Teams and SharePoint). This can all be permanently lost when these assets are processed once the employee leaves.

In the past, it was common for companies to simply reassign the employee’s Office 365 license, which meant losing all employee email, calendar, OneDrive and SharePoint data after 30 days. This is done in order to re-image workstations, but again causes a loss of all data. Quite simply, a re-image will delete the employee’s share drive folder to reclaim storage.

Today, everyone’s more aware of the potential value of ex-employee data, but even more important is the need to retain this data in case of future wrongful termination lawsuits.
Now, you’ve likely created policies and procedures to ensure the capture and protection of all departing employee data. Chances are, you’re keeping it for a set time period based on your local statute of limitations, just in case you ever need to review the data if a lawsuit was to be filed in the future.

One of our pharmaceutical customers experienced something similar to this, needing to capture and store images of ex-employees’ smartphones and workstations in order to produce them in the incident of legal or regulatory actions. With a data lake, they could search easily and navigate through the data if the need ever arose.

Picture the scene; a walled garden and a data lake

It sounds like a beautiful scene in the country, but a 'walled garden' eDiscovery data lake is actually something you could be using to save yourself time and reduce your risks of data loss or corruption.

With the right data management and eDiscovery software, a walled garden eDiscovery repository can be created that would act as a common legal department repository. This would remove the need to ship or move sensitive data around to your company's external counsel, or even opposing counsel, hence saving you that time and protection from loss.

Case study: DE Techs Inc v. Dell Inc lawsuit

Back in 2007, during the discovery phase of the DE Techs Inc. v Dell Inc. lawsuit, a motion was filed by the plaintiff's attorney complaining that Dell was making them search raw data stores for relevant case-specific content. They expected that Dell would follow the industry norm and perform the initial search and culling for them.

This was because many years ago, Dell made the decision to create a common eDiscovery repository where copies of all email and other documents were stored. This meant internal Dell attorneys and any plaintiff attorneys could access, search and review the results to their hearts' content – with the appropriate access controls and auditing, naturally.

Dell calculated that this process would save them huge amounts of money and resources, both in IT and legal personnel. In this case, the plaintiff's attorney suggested that Dell was making the employees conduct the company's eDiscovery. However, the Judge ruled in favor of Dell, stating;

"Providing access to a searchable database meets [the defendant's eDiscovery] requirement."

This was one of the earliest examples of a common eDiscovery repository, and a prelude to an eDiscovery data lake.

A trip to the (data) lakes just makes sense.

As more lawsuits are filed and eDiscovery datasets continue to get bigger, are held longer and continue to be a major target of cyber-criminals, storing them on-premise in department file sharing systems is no longer the best practice.

The cost of staying ahead of emerging cyber-threats is becoming more and more impossible. In fact, most companies are failing to even catch up, let alone put preventative measures in place. Instead of putting yourself through these expensive games of chase, you're hopefully starting to realize that taking advantage of the technological power of established public clouds is less risky, more secure, and less costly than trying to do it yourselves on-site.

Water you waiting for? Dive into a data lake!

Okay, the pun was bad, but stick with us here. We've got some questions you might want to ask yourself, or your colleagues, to help you focus on the real challenges you're facing today.

Questions to ask yourself or your GC:

1. Do you know where your eDiscovery datasets are currently stored?

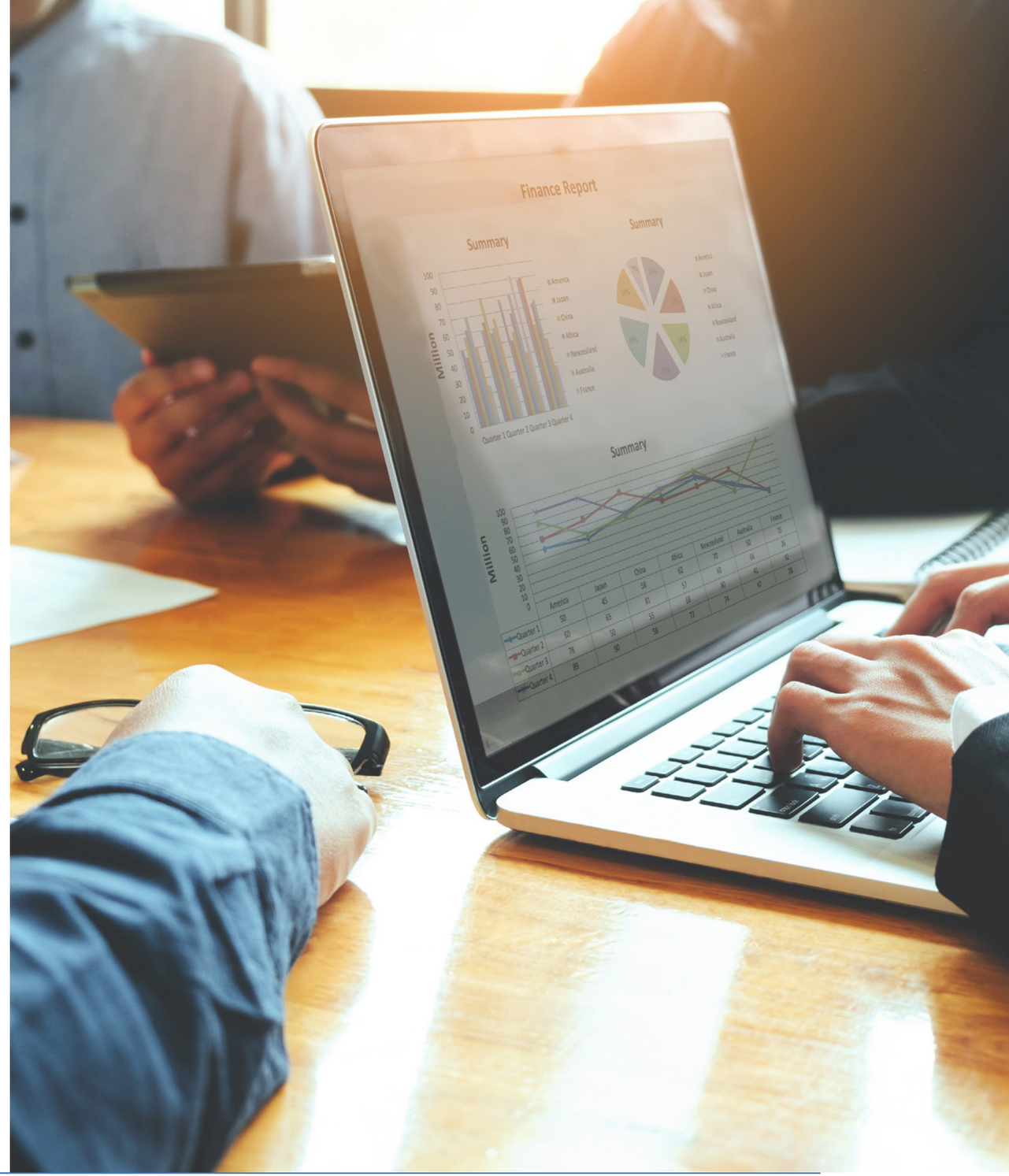
2. If eDiscovery datasets are stored on-premises, how do you place and guarantee litigation holds?

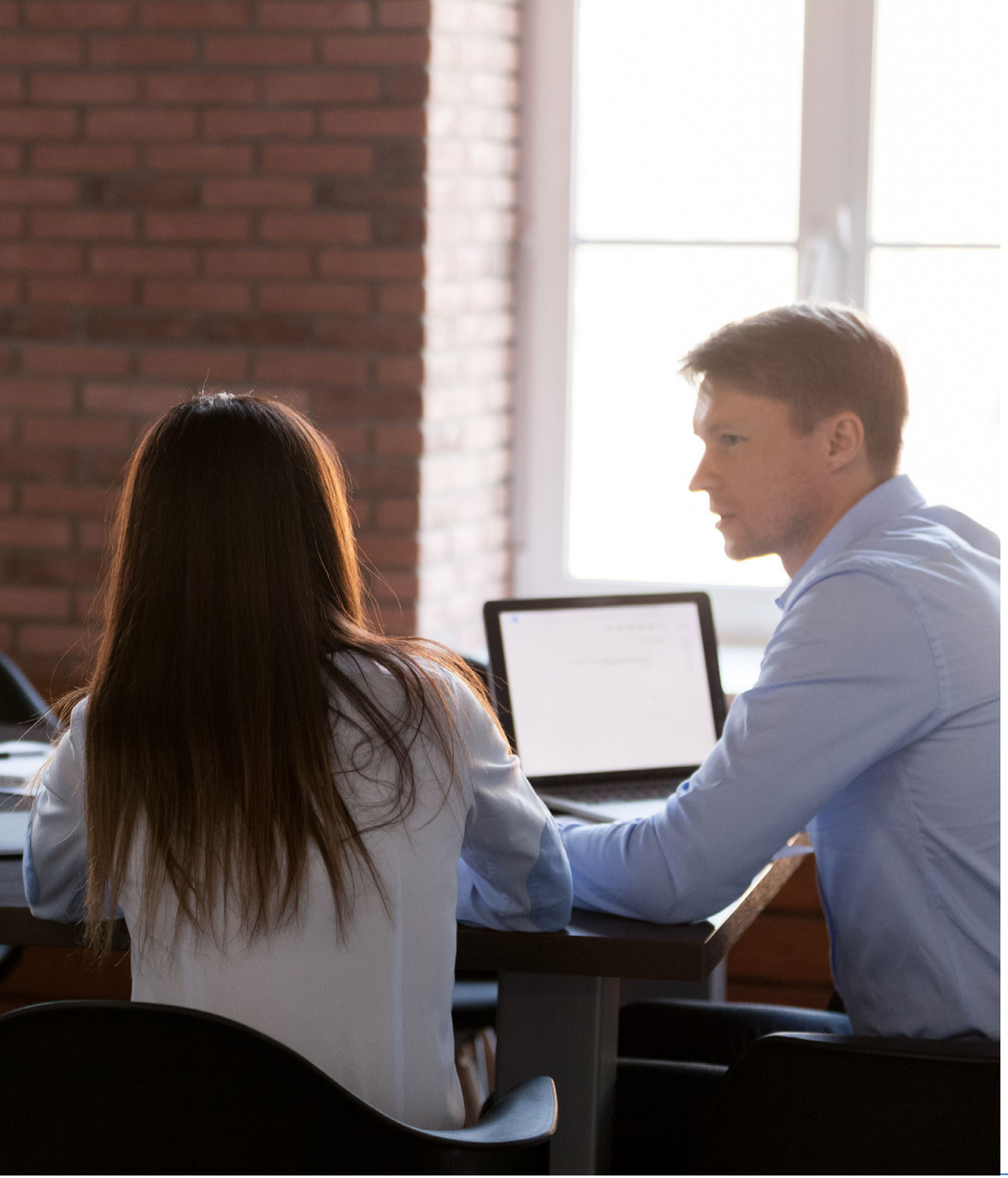
3. Does the legal department have a (regularly followed) process to delete aging, unneeded eDiscovery datasets?

4. Do you provide remote access to eDiscovery datasets so your attorneys can work when not in the office?

5. Do you provide access controls for specific eDiscovery datasets?

6. If you could store and actively manage your eDiscovery datasets in a secure department cloud, would you?





Questions to ask yourself or your CISO:

1. If you currently store the legal department's eDiscovery datasets on-premises, what types of additional security do you employ to protect this sensitive data from both internal and external threats?

2. Have you ever experienced a breach of your eDiscovery data repository…That you know of?

3. Would you be able to tell if your legal department's data repository was accessed inappropriately?

4. If breached, could you technically prove to a Judge that your data wasn't removed or altered?

5. If you could store and actively manage your legal department's eDiscovery datasets in a secure department cloud, would you?

6. Have you looked in to Azure or AWS for storing sensitive corporate information?

ARCHIVE360

I help our enterprise and government customers translate complicated eDiscovery, compliance, privacy, and records requirements into practical solutions rooted in modern technology. I'm available to answer any questions about how the cloud, analytics, and machine learning can help with automatic records classification, early case assessment, technology assisted review, GDPR, and emerging compliance regulations.

To schedule time with me, just send me an email:
bill.tolson@archive360.com

Bill Tolson | Vice President, Compliance & eDiscovery

Back to top