FILE REPUBLIC — 2024

Dark Data Report.

2024

FILE REPUBLIC

About File Republic.

We provide digital privacy and productivity solutions for law firms, allowing you to concentrate on your clients and cases. We aim to transform your old files from forgotten to effective and empower you with data. Whether it's legacy cases or ongoing matters, we provide storage, digitisation, and compliant, searchable access to all your data. With a single recoverable charge per matter, you gain complete control.

FILE REPUBLIC PRIVACY

Leveraging cutting-edge tools like AI and optical character recognition, we safeguard your digital client data with world-class privacy and compliance. Our robust information security measures protect Personally Identifiable Information (PII), streamline discovery processes, facilitate metadata structuring and document classification.

FILE REPUBLIC STORAGE

We offer comprehensive physical and digital file management solutions. Our services include free file audits, paper-to-digital conversion, and fast access to both physical and digital files.

FILE REPUBLIC PRODUCTIVITY

We provide powerful productivity solutions for professional, court-ready document bundling. Our platform offers features such as auto pagination, powerful searching capabilities, real-time collaboration, simple internal and external hyperlinking, and quick redaction tools.

Dark Data. Killing it before it kills your firm.

Over the last 12 months we have noticed more law firms questioning the unbillable hours spent on managing their client files to remain compliant with data privacy obligations.

Legal firms process a lot of data, but not all of it is easily accessible. Dark data is the unprotected, unused information sitting in warehouses and outdated digital systems. It affects your productivity, profitability, and case outcomes.

This report explores the high costs, compliance risks, cybersecurity threats, and wasted potential associated with dark data.

The good news? Active file management paired with incredible new technology such as Artificial Intelligence (AI) and Optical Character Recognition (OCR) technologies can turn this liability into a goldmine—one that empowers your team to achieve the best possible outcomes for your clients.

It's time to act. Assess your data practices, embrace new technology, and transform your dark data into a strategic asset.

This isn't about keeping up. It's about staying ahead.

REPORT BY
FILE REPUBLIC

What is **Dark Data**?

Think of your data like an iceberg. What's above the surface is the information you can readily access and use.

What's below the surface—potentially the bulk of your data—is not accessible. It sits unseen and unused, and your intellectual property (IP) goes to waste.

In the context of a legal firm, dark data includes both paper files stored in warehouses and offices and digital files buried within practice management systems.

Dark data poses many threats and challenges, one of the most critical being security. Dark data is not actively managed or monitored. It's more or less forgotten, and this means it's likely unprotected and not on the compliance radar.

For this reason, dark data management and cybersecurity are connected, but they aren't one and the same.

Cybersecurity is about protecting data from unauthorised access that leads to breaches and theft.

In contrast, dark data is the organisation, management, protection and usability of data. It's about formulating and upholding a data structure that enables team members to access and use it to win cases.

DEFINITION:
DARK DATA IS INFORMATION COLLECTED, PROCESSED, AND STORED DURING REGULAR BUSINESS ACTIVITIES BUT NOT ACTIVELY USED, PROTECTED OR ANALYSED.

How does it become **Dark**?

Data might start its journey in the light, but several missteps can cause it to go dark. Let's use a case file as an example.

Say an important case is filed away in a firm's records. It's not indexed or tagged in a way that allows for easy searching. The case's details—like the dates, case type, and involved parties—are not recorded in a searchable format.

Now, imagine a new case comes in that could benefit from the information within that document, the team is unable to find it. The employee who initially handled the document has left firm, taking with them the knowledge of its existence and relevance. Without proper indexing, the document remains buried in the archives, essentially invisible to the current team.

The data contained within this case has gone dark—it cannot be found or used to support the new case.

Here are some of the other reasons data can become dark:

01 **Awareness**

Employees are not aware of the data they receive or can access. They don't fully understand its value, don't store, manage, or use it proactively.

02 **Hidden data**

Metadata isn't considered despite the valuable information it holds, such as precise time and date stamps, device types, and geo-location.

03 **Data siloes**

Employees, teams, and departments store data independently, creating fragmented data siloes.

04 **Redundant, Obsolete, Trivial**

High volumes of data, that include multiple copies of the same data and outdated information.

Types of **Dark Data.**

Not all Dark Data looks the same. Overlooking any of it could cost you the opportunity to perform at your peak.

Types of dark data

Structured data includes data organised into defined fields, such as databases and spreadsheets. A customer relationship management (CRM) database is one example. Structured data may be easier to search and categorise, but it can still become dark.

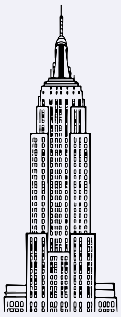
Unstructured data includes data that doesn't slot easily into defined fields. Examples include emails, PDFs, and chat logs. Teams or software will typically need to convert and organise this data before it can be used and analysed.

Semi-structured data contains some defined data fields but is not as searchable as structured data. Examples include invoices and graphs.

Metadata is data about data. It provides information about the characteristics of data, such as its structure, format, location, and context, enabling efficient organisation, retrieval, and management of information.

15%

OF THE GLOBAL DATASPHERE IS PROCESSED AND TAGGED TO MAKE IT USEFUL*



THE LEGAL INDUSTRY GENERATES 600 TIMES THE EMPIRE STATE BUILDING IN DATA EVERY YEAR

- https://www.seagate.com/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf

How does **Dark Data** affect you?

And, more importantly why, are law firms starting to worry about its implications for their reputation, results and regulatory risk?

Dark data drives up costs, puts your firm at risk of compliance issues, and turns your IP from an asset to a liability.

Let's examine how dark data impacts law firms and the resulting risks.

01 High overheads

Even digital data demands storage space, and that can be expensive. But perhaps the most pressing cost implication is dark data's impact on efficiency. Consider the time it takes for a lawyer to dig through unorganised files or search email inboxes for a specific insight or fact they need to support their case. Their time could be better spent preparing a compelling strategy, collecting expert testimony, or interacting with clients. This could translate into less successful case outcomes, affecting your firm's reputation and discouraging repeat business and referrals.

02 Compliance risks

Legal firms must comply with data storage, protection, and accessibility regulations. Poorly managed data can lead to non-compliance, resulting in hefty fines and legal issues. For example, you must redact personal identifiable information and adhere to the compliant destruction of files.

- https://www.oaic.gov.au/__data/assets/pdf_file/0018/1287/app-guidelines-chapter-11-v1.2.pdf

How does **Dark Data** affect you?

03 Client privacy

Almost one in five cyber attacks targets the legal sector, and dark data is more vulnerable to cybertheft. Why? Because malicious actors find it easier to access and steal unprotected data. In short, it's vulnerable.
Adding pressure to the situation, legal firms handle a lot of sensitive information. This includes verification of identity (VOI) documents, client details and case strategies. A data breach can have severe consequences, including loss of client trust and legal ramifications.

04 Under-utilised IP

Your firm generates a wealth of information that can support your team as they build and present cases. However, if they are guilty of data hoarding—saving everything "just in case"—they dilute the value of your data and make accessibility more complex.
If value-add data is obscured or inaccessible, staff cannot make the most of it. This results in missed opportunities and wasted IP.

05 Reputation damage

A firm’s reputation is crucial in the legal industry, and poor data management can harm how others perceive your credibility, professionalism, and integrity. Clients expect their information to be handled responsibly. Data breaches and non-compliance can lead to negative publicity and the loss of clients.

06 Un-billable hours

Managing dark data takes time—time that cannot be billed to clients. Lawyers and staff spend hours sorting through unstructured data, which reduces the time available for billable work. By streamlining data management, you can free up more time for billable activities, ultimately boosting your firm's profitability and productivity.

What happens when client data is compromised?

Sensitive client information is not always safe.

Take the CTS security breach as an example.
In November 2023, IT service provider CTS experienced a cyber attack, leading to significant disruptions and data leakage for around 80 law firms. The impacted conveyancing firms faced a major crisis—they were unable to complete transactions, leaving many clients in distress.

Deals fell through due to the outage. Removal companies had to be cancelled, and clients were frustrated by the violation of their privacy and lack of information and progress.
CTS acknowledged the incident and stated they were working "around the clock" to restore services. But despite their efforts, they could not provide a precise timeline for full restoration. This uncertainty further aggravated the situation for firms and their clients.
The CTS data breach is not an isolated incident. These events pose a serious threat to the legal sector's efforts to protect client information.

- Sixty per cent of identified data breaches in the legal sector are caused by insiders. By comparison, 40% of data breaches come from external threats, like malicious actors. Human error is the most common cause of breaches, accounting for 39% of incidents. This is closely followed by data being shared with the wrong person.

- In the UK, data breaches at legal firms compromised the information of 4.2 million people, affecting 6% of the nation's population. Nearly half of the breaches impacted customers, while 13% affected employees.

- The main types of data breached included basic personal information, economic and financial data, health data, and official documents.

Dark data is not exclusively about data breaches, but it is a crucial piece of the puzzle. If firms don't know what data they have or where it's stored, they can't protect it properly.

https://www.lawgazette.co.uk/news/dozens-of-conveyancing-firms-paralysed-by-cyber-attack/5118045.article

Are you ready?

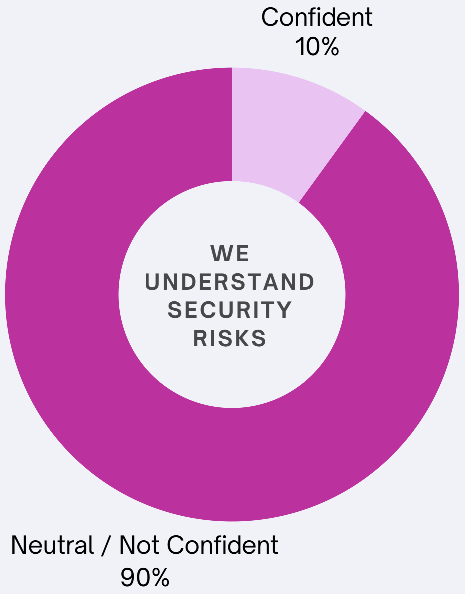
Our industry survey shows that law firms are aware of their duty, but lack the resources to effectively manage their data.

Our research indicates a startling gap in the legal industry. Firms don't understand the true nature or extent of security risks.

This makes their data even more vulnerable to breaches. Sensitive business and client data is vulnerable, and firms are left exposed to suffering irreparable financial and reputational damage.

In addition, those that aren't aware of or don't know how to meet data security standards are at a higher risk of non-compliance.

Firms with a strong reputation should be cautious of high-stakes dark data risks, while also recognising that active information management can deliver a competitive edge. Implementing a robust system of structured data can significantly enhance document review and e-discovery processes, providing those firms with a substantial competitive advantage.



40%

BELIEVE THEIR ORGANISATION DON’T REGULARLY REVIEW DATA PRIVACY AND SECURITY POLICIES.

FILE REPUBLIC CONSUMER INDUSTRY SURVEY 2023.

Challenges & Opportunities.

Firms rely too heavily on their practice management systems.

Many legal firms rely on their practice management systems. They overlook associated risks like limited backup solutions, where a single system failure could prevent business continuity. What's more, these systems often lack the tools needed for long-term file preservation or metadata extraction.

Opportunity: By diversifying backup solutions and integrating advanced tools for file preservation and metadata extraction, your firm can ensure seamless business continuity.

Firms fail to adapt and evolve alongside emerging risks.

Outdated data privacy practices can leave private and sensitive information vulnerable to cybercriminals. To counter new cyber threats, law firms must take additional precautions to safeguard their clients' data and meet new data privacy obligations. Regularly reviewing policies and monitoring retention practices are effective ways to demonstrate active compliance.

Opportunity: Adopt a proactive approach to data policy reviews and updates. This may include tweaking your governance strategy to minimise the production of dark data.

Firms don't invest in ongoing security training.

The low confidence levels indicated in our survey suggest a lack of staff training. This can lead to data being misplaced, improperly stored, or left unsecured. Additionally, employees who are not well-trained in data management are more likely to make mistakes. Keep in mind that almost 100% of cyber incidents involve some form of human error.

Opportunity: Invest in comprehensive training programs and automated systems to improve staff knowledge and confidence in data management.

How to prepare for **Dark Data.**

01 Find your Dark Data

The truth is that Dark Data can be hiding in multiple places due to legacy systems, storage facilities and siloed teams using different processes to manage information. In the short term it can be easy to dismiss this as a "job for later", but the longer this is left the harder it becomes when your firm grows larger and introduces more employees, clients and cases.

02 Assess your systems

Using the right tool for the right job remains crucial in the age of technology. While a practice management system enhances efficiency and effectiveness, it may not be suitable for managing clients' data privacy and your firm's compliance. Conducting a thorough assessment of your software and suppliers against information management criteria throughout the matter lifecycle is essential for addressing your dark data challenges.

03 Evaluate in-house vs outsourced expertise

Effective information management—encompassing governance, compliance monitoring, and knowledge management—can demand more than a full-time role. While managing an employee adds strain, it also brings intimate knowledge of your operations. Outsourcing, on the other hand, ensures focused, specialised expertise, a performance-based relationship, and access to resources that only scale can provide.

04 Set a periodic review

No matter how you choose to manage your data, it can't be a set-and-forget practice. Regular reviews are essential to control storage costs (both physical and cloud), stay current with compliance obligations, and identify new data privacy threats. Additionally, technological advancements offer new opportunities to leverage your intellectual property, which should be regularly explored to ensure you remain a leader in your field.

A focus on AI and OCR

To truly optimise your data use, you need both automation and consistency. This requires an airtight system for converting, classifying, and storing data in a compliant, encrypted, and easily searchable platform.

AI and OCR can help make these outcomes a reality. These technologies take your data utilisation to a whole new level, all with minimal manual effort and built-in compliance.

AI and dark data
AI simulates human processes like learning, reasoning, identification, and self-correction—faster and with greater accuracy. AI can:
•*Uncover hidden data*, extracting valuable information from high volumes of structured, unstructured, semi-structured, and meta-data data to equip your team with the insights they need.
•*Automate data classification and organisation*. It sifts through various data types and sources to categorise them and enable searchability.
•*Enhance data governance* by monitoring and managing data quality on a continuous basis. Data is then more easily stored securely and in compliance with regulatory standards.

OCR and dark data
OCR converts images of text—think scanned paper documents, photographs and PDFs—into a machine-readable format. It can:
•*Make every piece of data across your firm searchable*, regardless of its original format. That means your team can use simple search functionality to locate important facts or insights from all kinds of records and correspondence: letters, images, and more.
•*Find Personal Identifiable Information (PII)* hidden within images that can easily be overlooked by humans due to size, scale and visual complexity.
•*Improve accessibility*, eliminating data silos. Every person who needs access and is authorised can view relevant documents.
•*Reduce human errors* during data entry by automating the digitisation process.

The benefits of AI and OCR

> “ARTIFICIAL INTELLIGENCE WITH NO STRATEGY IS THE SAME AS INTELLIGENCE WITH NO STRATEGY “

Ben Hogan
Managing Director
File Republic

Save time and reduce costs

AI and OCR reduce the time spent on manual data organisation and the laborious process of digging through physical documents. Your firm's legal professionals can dedicate more time to case-critical tasks.
Plus, digital integration capabilities sync digital files into a system where metadata—like timestamps and GPS coordinates—are extracted and indexed to simplify e-discovery.

Improve data accuracy

Manual data entry and classification are prone to errors. AI and OCR improve accuracy by eliminating human error. They consistently and correctly classify documents based on their content, ensuring all data is properly categorised and easily retrievable.

OCR technology scans every part of your documents for personally identifiable information (PII), so no detail is missed.

More resources available for billable work

AI and OCR automate routine processes. Your team has more time to concentrate on valuable, profit-driving activities. This increases productivity and boosts job satisfaction.

Streamline case preparation

AI and OCR organise and store all documents correctly and in compliance with regulations, enabling best-practice data management. Digital files are converted to PDF format, making them universally accessible and interactive. All text is indexed and comprehensively searchable.

Team members can then quickly locate the information they need to prepare for client meetings and build winning strategies.

Meet data privacy obligations

AI can identify, classify, and tag sensitive information within dark data, significantly enhancing information security. This makes compliance with data privacy laws effortless, reducing the risk of accidental data breaches.

Additionally, reporting and monitoring tools are essential to stay ahead of compliance requirements, such as preserving files for the long term, sealing files to prevent tampering, and using encryption to protect the chain of evidence.

A final note from File Republic on Dark Data.

Even though we don’t want to admit this, data management is just about as un-glamorous as it gets. We don’t remember Harvey Specter worrying about compliant case destruction while he drove his vintage Aston Martin through New York City, or Jessica Pearson as she smashed another case, heck even Louis Litt didn’t mention it*.

But the truth is, if you’ve got someone else worrying about the ‘small’ things you can make sure you’re focused on winning cases and growing your firm. And, even though data management is a small part of running a successful practice it can become big enough to shut you down if you don’t do it right.

That’s why, for us, the small things are the BIG things. It’s why we obsess over the details, why we study compliance principles, why we build strategies to protect PII.

It’s also why, behind the simple idea that we provide ‘privacy and productivity to law firms’ there’s an incredible amount of software, computing power, global partnerships and constant innovation that to be honest is quite hard for us to articulate without verging on being downright dull (you wouldn’t want to be stuck with us at a party).
In a sense we’re a bit like home insurance, not fun to think about but you’ll wish you had us at your side when it matters most.

The good news is, the process is easy. We audit your current data, then play a complimentary role to your practice management system, keeping your client data private and secure, and help you with using your data for things like automated court bundling and risk reporting.

So, don’t invite us to your next party, but do invite us in for a demonstration. You won’t regret it.

*Apologies, we’re also huge fans of ‘Suits’.

Thank you.

Thank you for taking the time to read this report. If you have any questions or would like to discuss our findings further, please don't hesitate to reach out to us.

www.linkedin.com/company/file-republic

✉ sales@filerepublic.com

filerepublic.co.uk