LOKKER
POWERING PRIVACY.

Website Trackers, Pixels and Cookies

Key Actions Companies and Insurers Can Take in an Evolving Privacy Landscape

ONLINE DATA PRIVACY REPORT

March 2023

Millions of Hidden Pixels, Trackers, and Tags Are At The Heart of Web Privacy Lawsuits.

The recent burst of data privacy-related class-action lawsuits related to online trackers and social media pixels, plus the recent enforcement action against Sephora under the California Consumer Protection Act and action against GoodRx from the Federal Trade Commission under the Health Breach Notification Rule (both of which relate to their unauthorized collection of web data) indicates that we're beginning an expansion of US regulatory actions.
Getting control of website data privacy has never been more urgent for companies.

So, why is this happening?

According to HTTP Archive's latest "Annual State of the Web Report" (Sep. 2022), 94% of sites use at least one third party, and, on average, the top 1,000 websites use 53 third-party scripts including ads, analytics, CDNs, chatbots, video delivery services, content providers, and social media features. Introducing all this activity into millions of browser sessions has become mayhem for unauthorized data collection, theft, and exploitation. Thus, the customer's web browser is now a hotbed of cyber risk, exposing visitors to malware, theft of their private information, and violations of privacy laws.

Our Research

To understand the scale and scope of the risks that exist, we analyzed 170,000 websites around the world to quantify privacy risks. This report looks at the prominence of specific third-party risks across all the websites and then dives deeper into the frequency of risks on individual sites. Using our proprietary risk score, we have examined how much 'website privacy risk' is present at the individual website level across various industries and the S&P 500. In short, we found that far more companies, and therefore insurers, are at risk of regulatory actions and lawsuits related to data privacy.

Our research uncovered over 5.1 million data privacy risks lurking beneath the surface of company websites. The primary culprits are third-party JavaScript trackers, fingerprinters, data skimmers, and session replay scripts that, while some provide beneficial website features, they may also collect and share the visitors' information, often with unauthorized partners and without the website owner's knowledge.

LOKKER has identified over 5.1 million data privacy risks.

Companies and Cyber Insurers Are Both at Risk

When companies inadvertently share personal data with a host of third parties, they put themselves and their customers at risk. They risk losing millions of dollars in legal expenses, regulatory fines, and penalties. Not to mention losing consumer trust.

For companies with cyber insurance policies covering data privacy incidents, the insurer is the one to take on the burden of the financial repercussions ultimately. We've already seen several healthcare companies treat the data exposure caused by online trackers as a data breach, which can be a long, costly process of notifying customers and regulators and remedying the situation. Cyber underwriters and claims teams are now scrambling to revise their policy language (and rates) to address the growing data privacy risks in their books.

Our report dives into the results of our web privacy research and further outlines actions that companies and insurance agencies need to take to protect themselves in an evolving web privacy ecosystem.

METHODOLOGY

The 9 Web Privacy Risks You Need to Get Under Control

LOKKER has identified nine categories of third-party scripts that could wreak havoc with companies' web operations and compromise customers' privacy. LOKKER's research uniquely quantifies 'web privacy risk' by classifying third-party scripts into these categories. These risks also form the basis for the proprietary LOKKER Website Privacy Risk Score discussed on page 7.

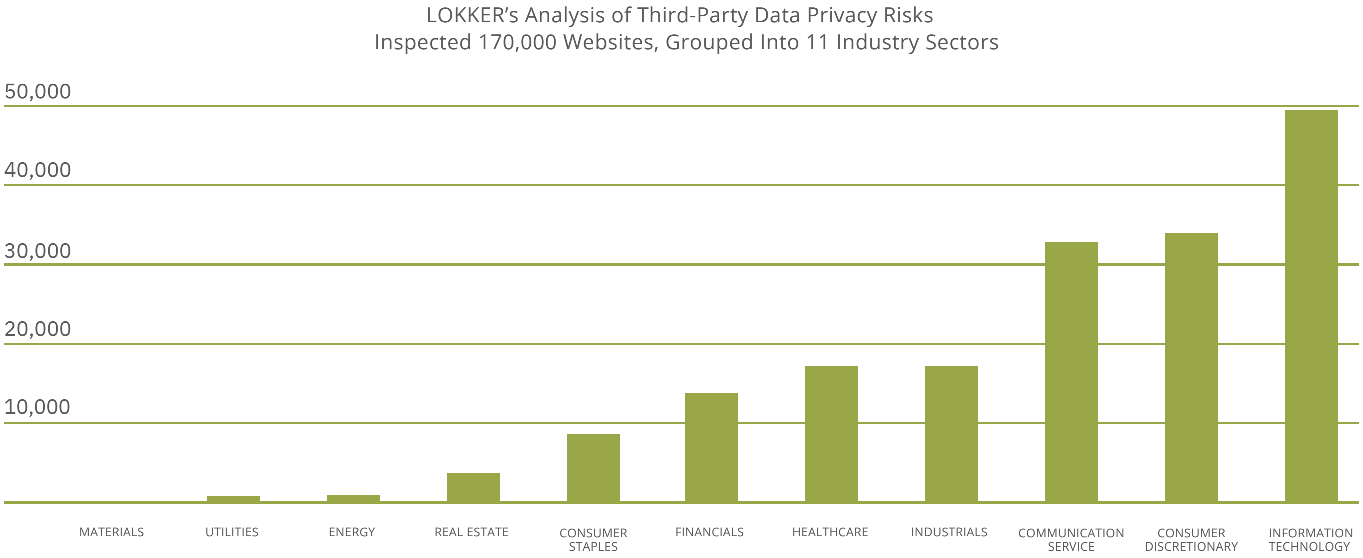
Web Privacy Risk Categories	Description
Malware	Malicious script that could deliver ransomware or keystroke logger that captures log-in credentials
Data skimming of sensitive data	Possible Magecart attack or unauthorized data sharing that leads to GDPR, CCPA or privacy policy violations
Trackers	JavaScript that includes some 'event' or 'action' script that collects information from the host site and transmits to a third-party. This is what often leads to unauthorized data sharing.
Cookies being set in each session	Third party cookies may not be authorized or violate privacy policy
Fingerprinting Scripts	Profiling of anonymous users that is used to link with personal identifiers by data brokers
Foreign domains making requests	Potential GDPR violation or risks by known nation-state actors
Session replay scripts recording activity	Sensitive data may be captured incidentally
Young domains serving JavaScript	Malicious scripts are often served by recently established domains
Bad SSL Certification	Domains serving scripts may not be secure and could indicate bad actors

Often, companies that deploy these scripts:

- Have no visibility to the extent of the data collection nor telemetry (sending data to site)
- Lose track of which scripts are still operating, beyond the term of a contract with that provider
- Have not configured features correctly and may be inadvertently collecting and sharing PII
- Cannot track the data that is shared with their legitimate partners that may also be shared with unauthorized 4th, 5th or Nth parties

Behind the Findings: Our Data Set

LOKKER's first edition of our online data privacy report (reported in March 2022) analyzed over 90,000 sites. For the second edition, LOKKER extended the research and analyzed privacy risk across 170,000 websites. Sources included Alexa, Cisco Umbrella, and DomCop, and categorized each site into one of 11 industry sectors.

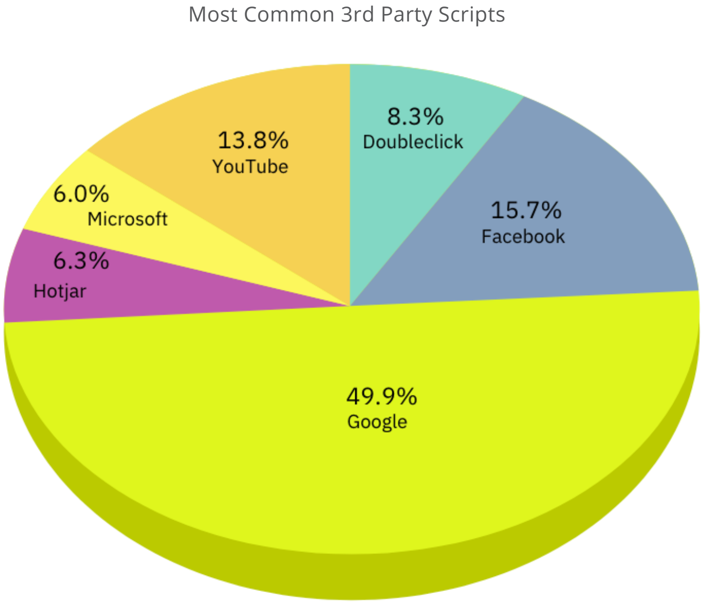


Majority of Third-Party Data Requests Are From Google, Facebook, and Microsoft

The chart below illustrates the 25 most common third-party scripts from the 170,000 sites LOKKER inspected. Unsurprisingly, the most requests were from the three online data giants, Google, Facebook, and Microsoft. These companies own multiple products like Google's analytics platform, ad services, APIs, and more that all have unique scripts, and sub-brands like Doubleclick ad network and Microsoft's LinkedIn platform.

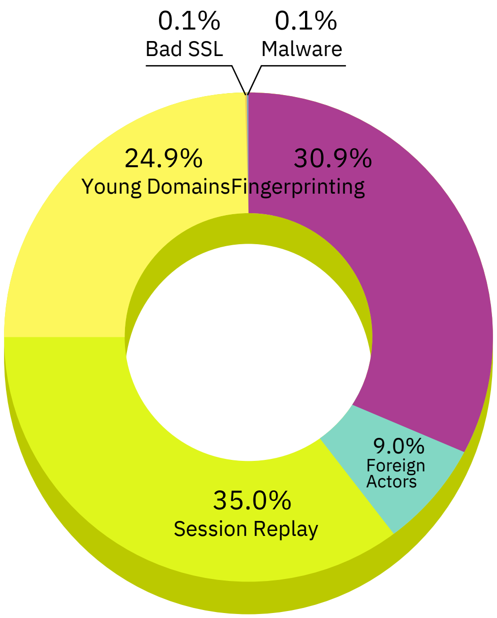
Session Replay Scripts

Rather unexpected was the prominence of session replay scripts. Most common was from Hotjar, seen on almost 40,000 sites. HotJar is one of several "session recording" tools websites use to record site visitor behavior to help site owners improve their user experience. However, if the tool is not configured correctly, it can record sensitive information. These session recorders are at the heart of a recent Pennsylvania data privacy lawsuit, claiming that several companies have violated wiretapping laws.



In addition to the risks that the JavaScript 'trackers' present to companies, LOKKER's analysis includes several other classifications of online data privacy threats, as follows:

Of the 170,000 Websites Scanned, The Following Data Privacy Threats Were Present



38,000+ Fingerprinting Scripts Lead to Widespread Consumer Profiling

As illustrated in the chart to the left, the proliferation of browser fingerprinting is significant and indicates that companies are becoming more sophisticated in creating alternatives to 'cookie data.' Fingerprinting scripts can capture details about the website visitor, including location, IP address, type of device, fonts installed, and other specifications of their computer and browser. While this information alone is not 'personally identifiable information,' it is used by savvy data brokers to create profiles ('fingerprints') that are continually enriched until the visitor can be identified.

Over 11,000 Scripts Originated in Russia, Belarus, China and Iran

While CDNs often serve various third-party JavaScript requests for optimized delivery, LOKKER can also identify the location of domains serving these scripts directly. As known nation-state actors often initiate many cyber attacks (ransomware, distributed denial-of-service), LOKKER has identified over 11,000 scripts originating from Russia, Belarus, China, and Iran, the vast majority of which are from Russia.

A recent class action lawsuit alleges Facebook used a web tracker to collect information from hospital website activities. According to the recent findings by The Markup, "A tracking tool installed on many hospitals' websites has been collecting patients' sensitive health information—including details about their medical conditions, prescriptions, and doctor's appointments—and sending it to Facebook." (The full Markup article can be found here.)

Top 6 Social Media Networks Are Harvesting Data from Education, Financial Services, And Healthcare Sites

Recent class action lawsuits filed in California alleged several healthcare organizations were sharing data with Facebook (Meta). LOKKER wanted to understand how widespread the issues were and uncovered that the "Meta Pixel" is not the only social media tracker deployed across hospitals, financial services, and educational sites.

Across the homepages of the **Fortune 1000** websites, Facebook trackers were identified on 46% of sites, Microsoft on 31%, Twitter on 21%, and Pinterest on 11%.

LOKKER's analysis of the **Healthcare industry** (in the US) found the following: Facebook on 40%, Microsoft on 13%, Twitter on 8% and Pinterest on 6% of over 5,000 hospital and healthcare services (more details below).

Of note, TikTok trackers were discovered on 5% of the sites analyzed. Regarding **Financial Services** sites in the US, LOKKER found that Facebook is on 36% of sites, Microsoft on 19%, and Twitter on 10%.

LOKKER's analysis of **Education** sites in the US (over 6,000 domains inspected) found that Facebook trackers are on 42% of sites, Microsoft on 15%, Twitter on 10%, and Snapchat and TikTok on 5% of sites.

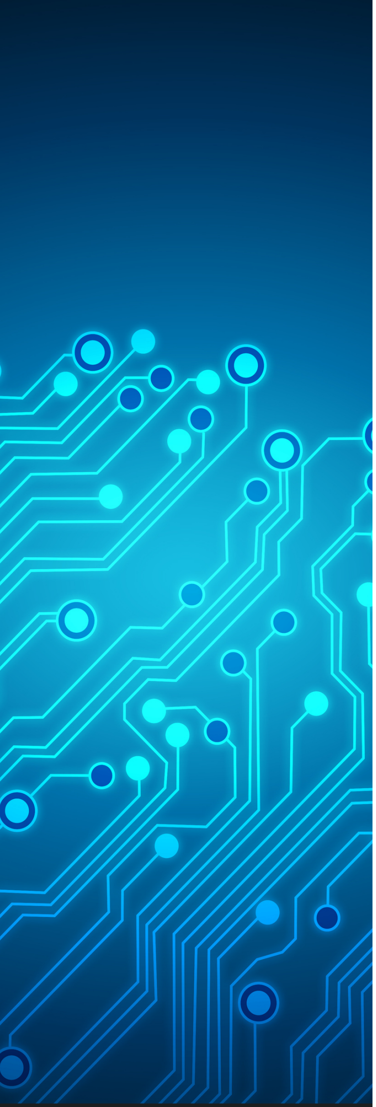
Presence of Oracle Tracker and Social Media Pixels on Inspected Sites

SECTOR	f	⊞	🐦	P	♪	👻	ORACLE
Fortune 1000	46%	31%	21%	11%	6%	6%	7%
Healthcare	40%	13%	8%	6%	5%	5%	8%
Financial Services	36%	19%	10%	2%	2%	2%	7%
Education	42%	15%	10%	2%	5%	5%	7%

Oracle is Fueling the Controversial Data Broker Ecosystem.

As the social media giants continue to collect and share web visitor data, this past August, Oracle was also named in a class action lawsuit filed in California, alleging that the company "has amassed detailed dossiers on some five billion people, accusing the company and its adtech and advertising subsidiaries of violating the privacy of the majority of the people on Earth."(Full article here).

LOKKER research uncovered that the Oracle "Bluekai" tracker (one of the elements under scrutiny in this lawsuit) and its "AddThis" cookie are widely distributed on sites throughout the world.



On Average, Sites Deploy 26 Cookies, Some Sites as Many as 322

LOKKER also analyzed the use of cookies across the 170,000 inspected websites. Cookies are essential to the data privacy conversation because cookies collect and store information (in the browser software) even after visitors leave a site. The risk to companies, however, is three-fold:

1. If a company is required to obtain consent from site visitors to issue cookies, is the consent tool listing all of the actual cookies?
2. If personal data is stored in the cookie, is it being accessed by unauthorized third parties?
3. If a visitor requests that data is not collected or stored, is that customer data removed from the company's records (or suppressed)?

Cookies are mainly used for three purposes:

Session management
Logins, shopping carts, game scores, or anything else the server should remember

Personalization
User preferences, themes, and other settings

Tracking
Recording and analyzing user behavior

The challenge for organizations regarding cookie management and data privacy compliance is to:

1. Ensure they have visibility to all of the cookies being placed into their customers' browser session
2. That only the data they wish to collect and share is used by authorized parties and in compliance with their stated privacy policy.

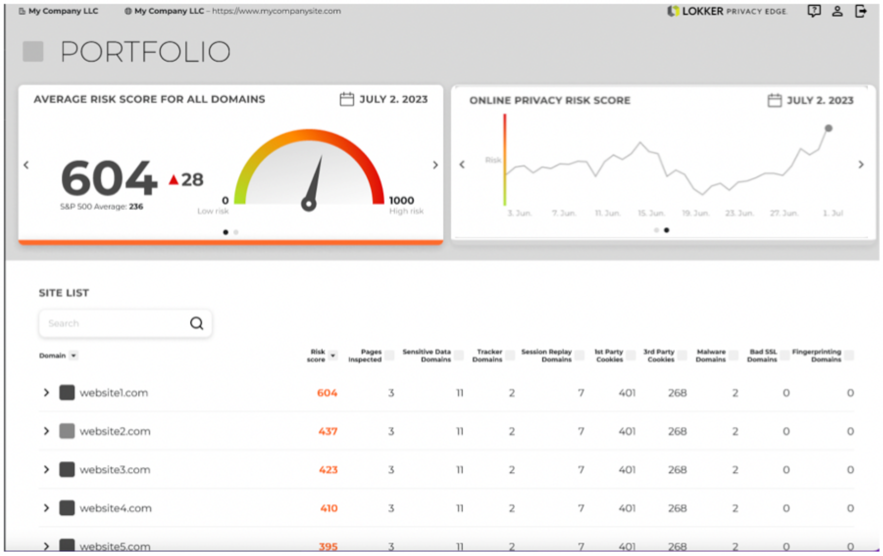
As you will see in the chart below, cookies also provide essential information that enables online ad targeting and tracking for companies like Google, Microsoft, Oracle, Casale Media, Bidswitch, etc. While some sites had as many as 322 cookies, we found, on average, that Fortune 1000 company homepages employ 26 cookies.

The 10 Most Common Cookies Found on Fortune 1000 Websites Are From These Domains:

Microsoft	.linkedin.com
Google	.doubleclick.net
AddThis (Oracle)	.addthis.com
Google	.youtube.com
Index Exchange	.casalemedia.com
Microsoft	.c.clarity.ms
Bidswitch	.bidswitch.net
AppNexus	.adnxs.com
Rubicon Project	.rubiconproject.com
Microsoft	.c.bing.com

Introducing The LOKKER Website Privacy Risk Score

It's important to know which risks are present on a website and the context in which they exist. That's why we developed the LOKKER Website Privacy Risk Score. The score is based on a proprietary calculation of the privacy risks present on a website, like trackers, cookies, and other third-party Javascript (which are further discussed on page 3). Each web page is scored individually, and then an average is taken to find the privacy risk score for the website. The score has adjusted weighting for the various risk types based on the severity, frequency, and placement of specific web privacy risks providing a more complete risk picture.



Domain	Risk score	Pages inspected	Sensitive Data Domains	Tracker Domains	Session Replay Domains	1st Party Cookies	3rd Party Cookies	Malware Domains	Bad SSL Domains	Fingerprinting Domains
website1.com	604	3	11	2	7	401	268	2	0	0
website2.com	437	3	11	2	7	401	268	2	0	0
website3.com	423	3	11	2	7	401	268	2	0	0
website4.com	410	3	11	2	7	401	268	2	0	0
website5.com	395	3	11	2	7	401	268	2	0	0

Website Privacy Risk Score: Research Findings

To better illustrate the usefulness of the risk score, we scanned and scored 1,000 of the largest companies in the data-sensitive healthcare industry. A few interesting trends and insights came out of our research:

Trend 1: High Risk Score Prompts Further Investigation

Mental Health Website

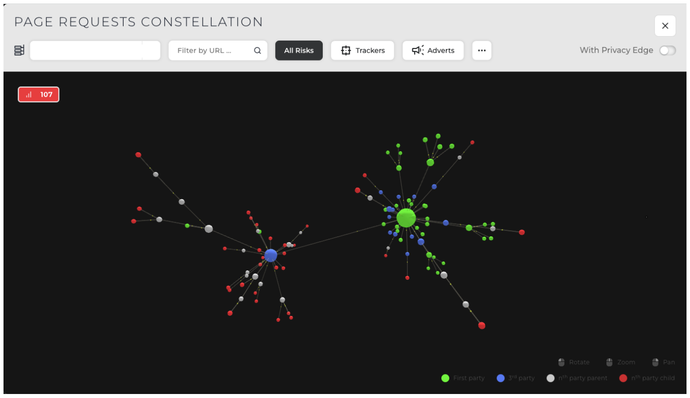
One mental health provider had a high overall risk score, prompting further investigation. We found the site used two session replay scripts: Hotjar and Crazy Egg, and had trackers including those from Oracle's Bluekai, Twitter, LinkedIn, and Facebook (Meta). The Meta pixel was present on sensitive health pages - for instance, those with information on trauma, eating disorders, and depression. What's more, while a user was browsing these pages, they were sending back data to Facebook with button clicks within the page - for instance, to enlarge photos of the residential care facility or to understand the therapy approach. For example, on a page that was to refer patients for residential care, once you filled out the patient details and clicked the 'submit' button, they sent this event to Facebook as a 'SubscribedButtonClick,' along with the first party cookie (which could be used to link a users activity to their Facebook account).

Trend 2: Problematic Web Pages Get Brought to the Surface

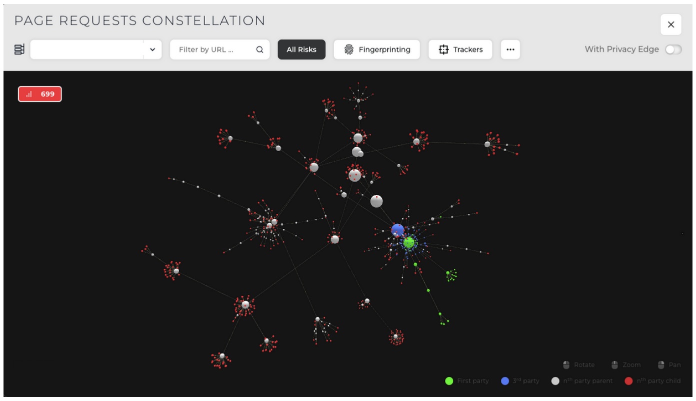
Large Hospital Provider Has High Risk Web Page

We found that one large healthcare provider had a low website privacy risk score, but an individual webpage had the maximum risk score possible. It was the only individual page on the site that ranked this high. The page had three session replay tools and many fingerprinters, including adware that causes intrusive popups and offers for the user. But it points to a more significant issue: just one page could lead to the breach of a visitors' data or potentially get a company sued.

It's easy to conceive how this could happen; in large organizations, multiple people are often responsible for different parts of the website, which could lead to inconsistencies in the website experience and a lack of oversight. For example, this individual webpage had embedded videos, third-party advertisements, a chatbot, a comments section, newsletter form asking for health symptoms/diagnoses; all of these are third-party tools that collect and share visitors' data. We've included a visual showing all the third-party JavaScript on both this site's homepage and the high-risk page to visualize the discrepancy further.



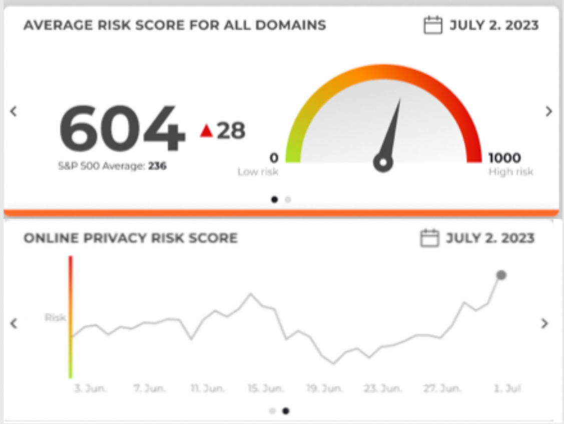
Above is an example scan of the hospital homepage. The green represents first-party data sharing, meaning it's an entity the hospital owns. The blue represents a third-party provider receiving data from the website. The gray and red dots are other parties with which the third party shares data, often without the first party's knowledge. There are 107 entities exchanging data on the homepage.

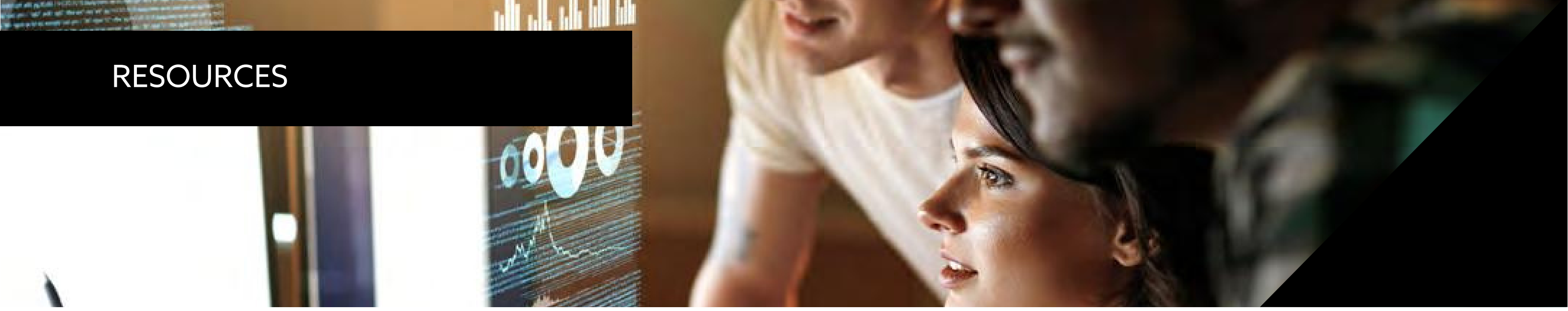


Compared to the homepage scan, you can see that this graphic is overwhelmed with gray and red dots - downstream data partners. There are 699 entities receiving data from the website, compared to 107 from the homepage.

The Website Privacy Risk Score: Providing Risk Visibility Across Portfolios and Websites

As illustrated in the examples above, the website privacy risk score is another measure for quantifying the totality of risks across an individual website and the individual pages on a website. For cyber insurers or other providers who provide privacy services to multiple organizations, the web privacy risk score is a valuable tool to identify high-risk websites and problematic webpages across their entire portfolio, helping to focus attention where it is most urgent. In addition, the privacy risk score is a valuable tool for assessing privacy risk during the underwriting process, fixing threats as part of an incident response program, or managing a proactive cyber management program. Finally, the score is a valuable metric to track over time and report to clients.





RESOURCES

Prepare for 2023 Privacy Regulations

Given the recent class action lawsuits, the growing consumer concern about personal data privacy, and the imminent data privacy laws going into effect in 2023 in California, Virginia, Utah, Colorado, and Connecticut, there is increasing pressure on organizations to take control of data privacy risks on their websites.

For reference, new data privacy laws going into effect
in 2023:

California California Consumer Privacy Rights Act
(CPRA) Jan 1, 2023
Virginia Virginia Consumer Data Protection Act
(VCDPA) Jan 1, 2023
Colorado Colorado Privacy Act
July 1, 2023
Utah Utah Consumer Privacy Act
December, 2023
Connecticut Connecticut Data Privacy Act
July 1, 2023

Additional data privacy laws referenced in legal actions:

Health Insurance Portability and Accountability Act (HIPAA)- FTC fined GoodRx under the HIPAA Health Breach Notification Rule. Several lawsuits were brought under HIPAA against hospital providers regarding the improper use of the Meta pixel

Wiretapping Laws - State wiretapping laws have been used to bring legal action against session replay providers and the companies deploying the tools on their sites.

Video Protection Privacy Act (VPPA) - Lawsuits have exploded recently alleging the collection and sharing of protected video consumption information.

Resources for Companies

LOKKER recognizes that organizations have an immense responsibility to protect their customers and their company. In addition to constantly fending off threats from malicious actors, security, marketing, and privacy executives have the additional challenge of complying with data privacy laws.

There is a strong ecosystem of attorneys, technology advisors, and software providers that can help organizations prepare for and mitigate data privacy risks.

The International Association of Privacy Professionals is a great resource (IAPP.org). The team at The Markup also provides timely, valuable research and insights. Helpful information is also available at The Electronic Frontier Foundation and The Center for Humane Technology.

For more information on the findings in this report, or to learn more about LOKKER's data privacy solutions, contact:

Jeremy Barnett
Chief Commercial Officer
jeremy@lokker.com
LOKKER.com

About this Report

LOKKER provides this online data privacy report twice each year to help organizations identify emerging data privacy threats and support their efforts to make web experiences safer for their customers. In addition to the research in the report, LOKKER provides its Privacy Edge™ software to organizations to continuously monitor, alert and block threats in the browser.