Thomson Reuters collected and leaked at least 3TB of sensitive data

Published: 2022-10-27 · Archived: 2026-04-15 02:03:03 UTC

Thomson Reuters, a multinational media conglomerate, left an open database with sensitive customer and corporate data, including third-party server passwords in plaintext format. Attackers could use the details for a supply-chain attack.

- Media giant with $6.35 billion in revenue left at least three of its databases open
- At least 3TB of sensitive data exposed including Thomson Reuters plaintext passwords to third-party servers
- The data company collects is a treasure trove for threat actors, likely worth millions of dollars on underground criminal forums
- The company has immediately fixed the issue, and started notifying their customers
- Thomson Reuters downplayed the issue, saying it affects only a "small subset of Thomson Reuters Global Trade customers"
- The dataset was open for several days – malicious bots are capable of discovering instances within mere hours
- Threat actors could use the leak for attacks, from social engineering attacks to ransomware

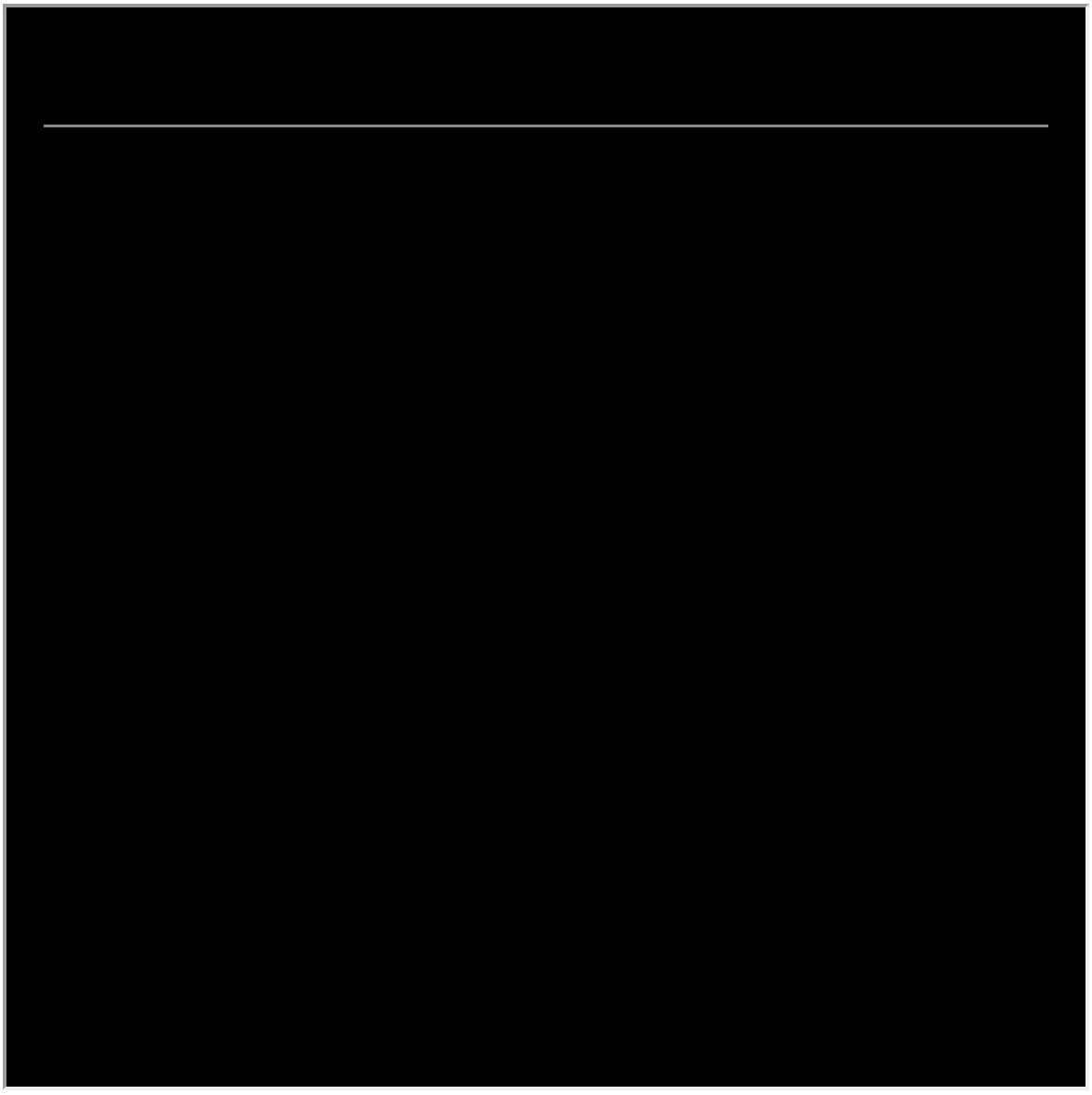
The Cybernews research team found that Thomson Reuters left at least three of its databases accessible for anyone to look at. One of the open instances, the 3TB public-facing ElasticSearch database, contains a trove of sensitive, up-to-date information from across the company's platforms. The company recognized the issue and fixed it immediately.

Thomson Reuters provides customers with products such as the business-to-business media tool Reuters Connect, legal research service and database Westlaw, the tax automation system ONESOURCE, online research suite of editorial and source materials Checkpoint, and other tools.

The size of the open database the team discovered corresponds with the company using ElasticSearch, a data storage favored by enterprises dealing with extensive, constantly updated volumes of data.

The naming of ElasticSearch indices inside the Thomson Reuters server suggests that the open instance was used as a logging server to collect vast amounts of data gathered through user-client interaction. In other words, the company collected and exposed thousands of gigabytes of data that Cybernews researchers believe would be worth millions of dollars on underground criminal forums because of the potential access it could give to other systems.

Meanwhile, Thomson Reuters claims that out of three misconfigured servers the team informed the company about, two were designed to be publicly accessible. The third server was a non-production server meant for "application logs from the pre-production/implementation environment."



The leaked data

Time stamps on data samples reviewed by the team indicate that the information was logged recently, with some pieces of data as recent as October 26. According to the researchers, the logs in the open database contain sensitive information and could lead to supply-chain attacks if accessed by threat actors.

For example, the open dataset held access credentials to third-party servers. The details were held in plaintext format, visible to anyone crawling through the open instance. According to Mantas Sasnauskas, the Head of Security Research at Cybernews, this type of information would allow threat actors to gain an initial foothold in the systems used by companies working with Thomson Reuters.

“ElasticSearch is a very common and widely used data storage and is prone to misconfigurations, which makes it accessible to anyone. This instance left sensitive data open and was already indexed via popular IoT [internet of things] search engines. This provides a large attack surface for malicious actors to exploit not only internal systems but a way for supply chain attacks to get through. A simple human error can lead to devastating attacks, from data exfiltration to ransomware,” Sasnauskas said.


Thomson Reuters data leak

Examples of passwords/credentials to a third party server (top) and connection string logs (below) on the database. Image by Cybernews.

The team also found the open instance to contain login and password reset logs. While these don't expose either old or new passwords, the logs show the account holder's email address, and the exact time the password change query was sent can be seen.

Another piece of sensitive information includes SQL (structured query language) logs that show what information Thomson Reuters clients were looking for. The records also include what information the query brought back.

That includes documents with corporate and legal information about specific businesses or individuals. For instance, an employee of a company based in the US was looking for information about an organization in Russia using Thomson Reuters services, only to find out that its board members were under US sanctions over their role in the invasion of Ukraine.

The team has also discovered that the open database included an internal screening of other platforms such as YouTube, Thomson Reuters clients' access logs, and connection strings to other databases. The exposure of connection strings is particularly dangerous because the company's internal network elements are exposed, enabling threat actors' lateral movement and pivoting through Reuter Thomson's internal systems.

There is a high chance the open instance included much more sensitive data since the database holds more than 6.9 million unique logs that take up over 3TB of server disk. The team claims that it is impossible to know the full extent of how big the dataset actually is without crossing the ethical boundaries within which researchers operate.

> *"This instance left sensitive data open and was already indexed via popular IoT search engines. This provides a large attack surface for malicious actors to exploit not only internal systems but a way for supply chain attacks to get through,"*
>
> *Sasnauskas said.*

The company's investigation

The team contacted Thomson Reuters upon discovering the leaking database, and the company took down the open instance immediately.

"Upon notification we immediately investigated the findings provided by Cybernews regarding the three potentially misconfigured servers," a Thomson Reuters representative told Cybernews.

According to the company, two of the servers were designed to be publicly accessible, while the third is a non-production server related to one of Thomson Reuters products, ONESOURCE Global Trade Product. The tool allows users to "manage export/import, sanctions screening, and other trade controls activities and related filings."

"This non-production server only houses application logs from the pre-production/implementation environment of that product and is only associated with a small subset of Thomson Reuters Global Trade customers," the company explained.

Non-production servers usually don’t hold application data. However, that does not mean that the details stored there are less sensitive.

“The open instance resembles a development server which can consist of an entire infrastructure and usually holds more sensitive client activity and data,” Sasnauskas said.


Thomson Reuters leak

Thomson Reuters says that the now-closed server only captures data generated through user actions within the pre-production and implementation environment.

“The server contains the information needed to operationally support the platform,” the company's representative explained.

However, it’s hard to tell whether all details stored on the instance were necessary to support the platform’s operations. Either way, even if all of the data was essential, that doesn’t make it less sensitive if leaked.

“Information stored on the server is extremely sensitive. Cases like these raise questions about corporate data collection practices. The ramifications of a data leak of such scale are worrying to say the least,” Sasnauskas explained.

The company launched an investigation to get the root of the problem. The leading theory so far is that an “isolated error in the product environment resulted in the inadvertent misconfiguration of the non-production environment.”

Thomson Reuters said it had begun notifying affected customers.

Significant impact

Researchers believe that any loss of information on the dataset could not only harm Thomson Reuters and its clients but also be detrimental to the public interest.

For example, the open database was leaking some individuals’ and organizations’ sensitive screening and compliance data. Accessible data from the public-facing Thomson Reuters database could have tipped off entities that would like their wrongdoing kept in the dark.

According to Martynas Vareikis, Information Security Researcher at Cybernews, threat actors could use the email addresses exposed in the dataset to carry out phishing attacks. Attackers could impersonate Thomson Reuters and send the company’s customers fake invoices.

> *“Information stored on the server is extremely sensitive. Cases like these raise questions about corporate data collection practices. The ramifications of a data leak of such scale are worrying to say the least,”*
>
> *Sasnauskas explained.*

“Having more details always helps malicious actors. Knowing the victims are Thomson Reuters clients allows for a targeted campaign. That’s especially true if Thomson Reuters clients used a non-public business email address to register with the company. Invoices infected with malware could cause huge losses for the clients if they were attacked by ransomware gangs,” Vareikis explained.

According to Sasnauskas there are numerous ways attackers could use the leaked details to harm the company itself. He claims that access to log files and the instance could enable malicious actors to leak sensitive information, extort the business, and gain knowledge about the internal networks, systems, and services in use.

“Attackers could pivot and move laterally in systems, and cause a plethora of malicious actions such as sell access to brokers or ransomware affiliates and launch sophisticated attacks, possibly including ransomware,” Sasnauskas said.

Why did it happen?

A thorough inspection of the SSL (secure sockets layer) certificate of the accessible web server, DNS (domain name system) data, and information on the ElasticSearch instance itself allowed the team to confirm that the open database belongs to the Thomson Reuters Corporation. The server had been left accessible since October 21.

IoT search engines did not show any results for the Thomson Reuters instance before that day. Since the web space is filled with bots and scripts hunting for open databases, it is doubtful the database was accessible to the public before.

According to Vareikis, the likeliest cause for the dataset to suddenly appear online is a misconfiguration error.

“We believe that it was caused by a misconfiguration on the AWS Elastic Load Balancing service, which followed different rules that weren’t configured to fully cover access control rules, which led to the service being exposed to the public,” Vareikis explained.

Exposed in the past?

Thomson Reuters security principles laid down in a whitepaper published last year claim the company’s secure configuration is created and deployed according to best practices.

However, digging through historical data from IoT search engines, researchers discovered that some of Thomson Reuters’ configuration and system environment files were exposed last year. Some of the files that appear on IoT search engines are still exposed to this day.

> *“This non-production server only houses application logs from the pre-production/implementation environment of that product and is only associated with a small subset of Thomson Reuters Global Trade customers,”*
>
> *the company explained.*

The company’s security principles also say that it performs automated and centralized logging to provide real-time alerting. However, the open dataset was accessible to the public for several days.

“It takes less than a few hours for an open server to be crawling with bots. Meanwhile, the data shows that the instance was open for more than three straight days. It begs the question of whether real-time alerting is necessary if there is no one to review the alerts,” Vareikis said.

Avoid at all cost

To deflect the impact that a lack of oversight might cause, it is highly recommended to avoid logging personally identifiable information (PII), such as corporate emails, in the same dataset with queries and other interactions.

Due to the sensitive nature of the requests that businesses might be using investigative tools for, exposing corporate inquiries may threaten to reveal company secrets, causing severe financial damage if made public.

“Even though the company encrypted communication with the server in SSL format, all security measures would fail with an introduction of a simple human error. There should always be measures mitigating these risks so that data does not get into malicious hands. Security practices, we see here, were not what you’d expect from a business as big as Thomson Reuters,” Sasnauskas said.

Companies should also avoid storing passwords in plaintext format. Even if databases are not public-facing, there are dangers of exposure. Credential theft and privilege-escalation attacks could allow malign actors to penetrate corporate databases, leaving passwords in plaintext format immediately exposed.

“Plaintext passwords to third-party servers stored in the open database should have been hashed with strong algorithms. That’s because even a strong password is obsolete once a database where it’s stored in plaintext is exposed,” Sasnauskas explained.

Source: https://cybernews.com/security/thomson-reuters-leaked-terabytes-sensitive-data/