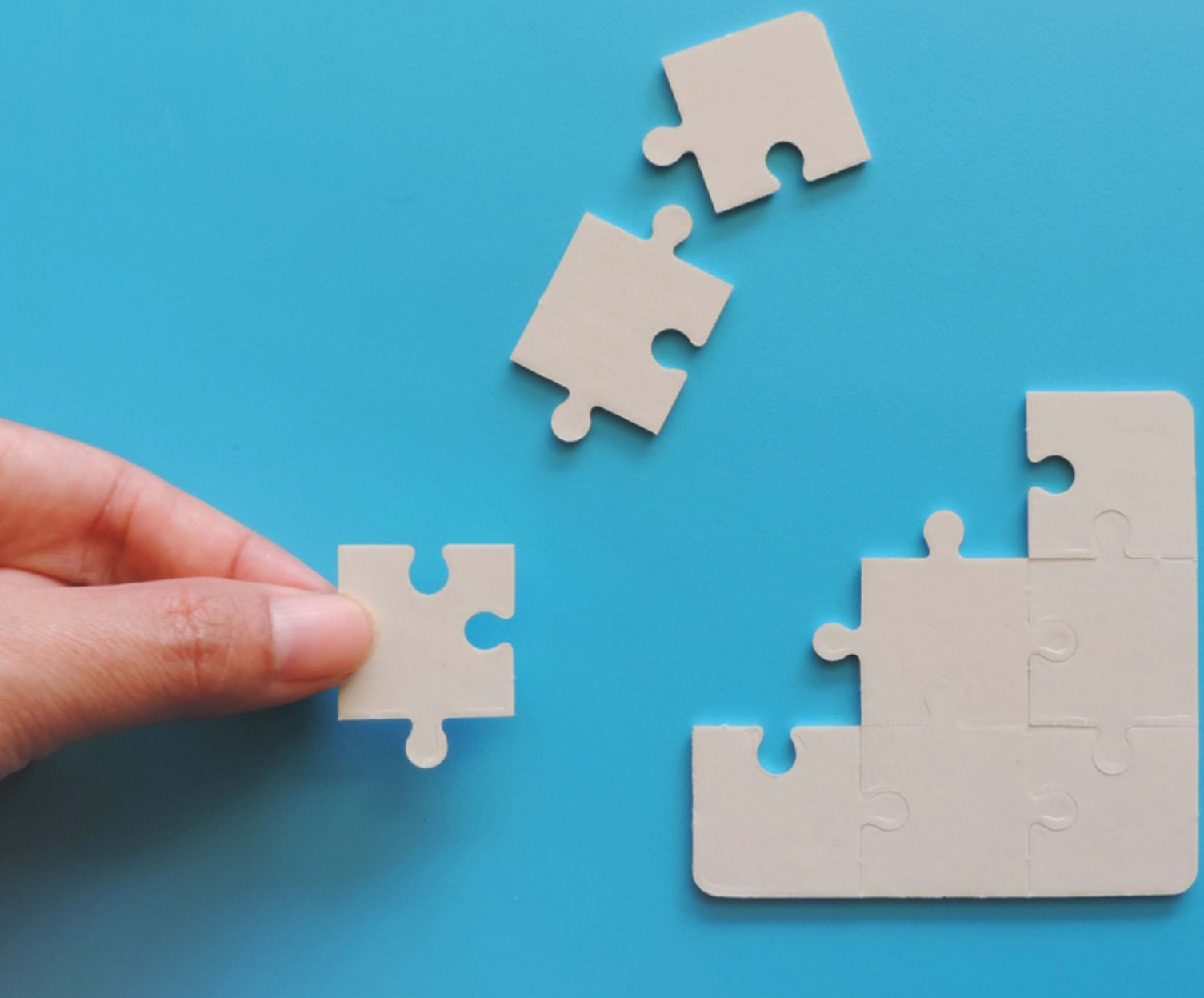
WhoisXMLAPI
The Who Behind Domain, IP & Cyber Threat Intelligence

WHITE PAPER

WHAT DO YOU PAY FOR WHEN BUYING COMMERCIAL INTERNET INTELLIGENCE DATA

WHEN FREE AND INTERNAL DATA FEEDS AREN'T ENOUGH, AND WHEN IS IT BETTER TO GO WITH PREMIUM INTELLIGENCE?



CONTENTS

Introduction 3

The Limitations of Free Internet Intelligence Sources 4

The Pros and Cons of Building Internal Internet Intelligence Feeds 8

The Benefits of Commercial Internet Intelligence Sources 11

Why Choose WhoisXML API Internet Intelligence Data 14

Final Thoughts 16

About Us 17

INTRODUCTION

Developing an effective cybersecurity product requires access to a broad range of Internet intelligence data — both real-time and historical — for a variety of tasks, from asset discovery to providing additional threat context.

Assessing domain legitimacy and threat level requires detailed WHOIS information. Botnet detection and infrastructure mapping rely on active DNS lookups, while passive DNS data is one of the backbones of malware campaign attribution and threat intelligence. Spam and phishing prevention rely heavily on data about newly registered domains (NRDs).

And there's always additional data that can add context, such as SSL certificates streams that can be used for better fraud prevention or IP netblocks data that many rely on to help with threat infrastructure attribution.

High-quality, reliable data sources directly influence the performance of cybersecurity solutions: the higher the quality and coverage of this data, the more reliable a product that uses it becomes and the fewer false negatives it will generate.

Some Internet infrastructure data is available for free, while other data sources are strictly commercial. So while it's possible to build and maintain in-house Internet Intelligence sources, doing so often involves significant technical complexity and long-term maintenance efforts that can be difficult to justify.

This white paper examines the strengths and limitations of free, commercial, and in-house data sources required to build cybersecurity products.

THE LIMITATIONS OF FREE INTERNET INTELLIGENCE SOURCES

It may seem that most information – such as DNS and WHOIS – is easily available for free via command line prompts. The `whois` or `nslookup` commands provide plenty of data to work with that could even be sufficient for the task during the stages of prototyping and early product development. But there are many limitations to them that in most cases prevent commercial cybersecurity products from relying on them, as well as other free data sources.

Request throttling

The primary problem a cybersecurity product that tries to rely on these commands would face is throttling: registrars and DNS servers often impose rate limits and start throttling or blocking requests if they detect an unusual number of queries coming from one source. This makes it hard to perform `whois` or `nslookup` commands at the scale needed for a cybersecurity product such as a threat intelligence platform or external attack surface management platform to function properly.

Lookup speed

Another problem is that lookups take time due to the absence of the locally cached data, possibly slowing the product down and making users wait. For some use cases, an ideal scenario would be to have a WHOIS or DNS database at hand that the product can query directly to speed up the process. However, free WHOIS or DNS intelligence doesn't come in the form of comprehensive databases – if you want one, your options are either to use a commercial solution or try building your own.

Protocol changes

In the case of WHOIS data, the recent (and continuing) migration to the RDAP protocol adds another layer of complexity to data collection and processing. While some registries (mostly in charge of gTLDs) have migrated to RDAP successfully and are disabling the WHOIS port 43, others (mostly in charge of ccTLDs) haven't properly implemented RDAP yet, or are not expected to implement it, and still rely exclusively on WHOIS.

That means that to get WHOIS information from some domains, you'll need to use WHOIS, but for others, it would be RDAP. So, you'll need to maintain a database that covers which query to use for which domain names. However, the situation is dynamically changing, with more and more registries offering RDAP and disabling WHOIS, so you'll need to keep the database's sources up to date.

In contrast, commercial providers such as WhoisXML API continuously monitor these changes and implement automatic fallback and other mechanisms that make sure you get the data no matter what changes occur on the registries' or registrars' side.

Lack of historical data

In many cases, cybersecurity products benefit from historical WHOIS or DNS data. It allows them to uncover domain relationships, track infrastructure ownership changes over time, attribute malicious activity to specific threat actors, and detect patterns that would be invisible with only real-time data.

However, neither WHOIS nor DNS protocols were designed to provide historical data. The ways to get it are either to use a commercial provider that has already collected this data for years or to begin collecting this data yourself. In the latter case, the historical depth would be limited with the date and time you start to collect the information.

Lack of global perspective

When it comes to DNS, running an `nslookup` command only shows the view from a specific resolver or geographic location. But a domain might behave differently across the world, as DNS responses can vary based on where and how a query is made — due to CDNs, load balancers, geofencing, or malicious intent. Relying on a single, local view can result in missing alternative IP addresses, hidden threats, or regional behaviors.

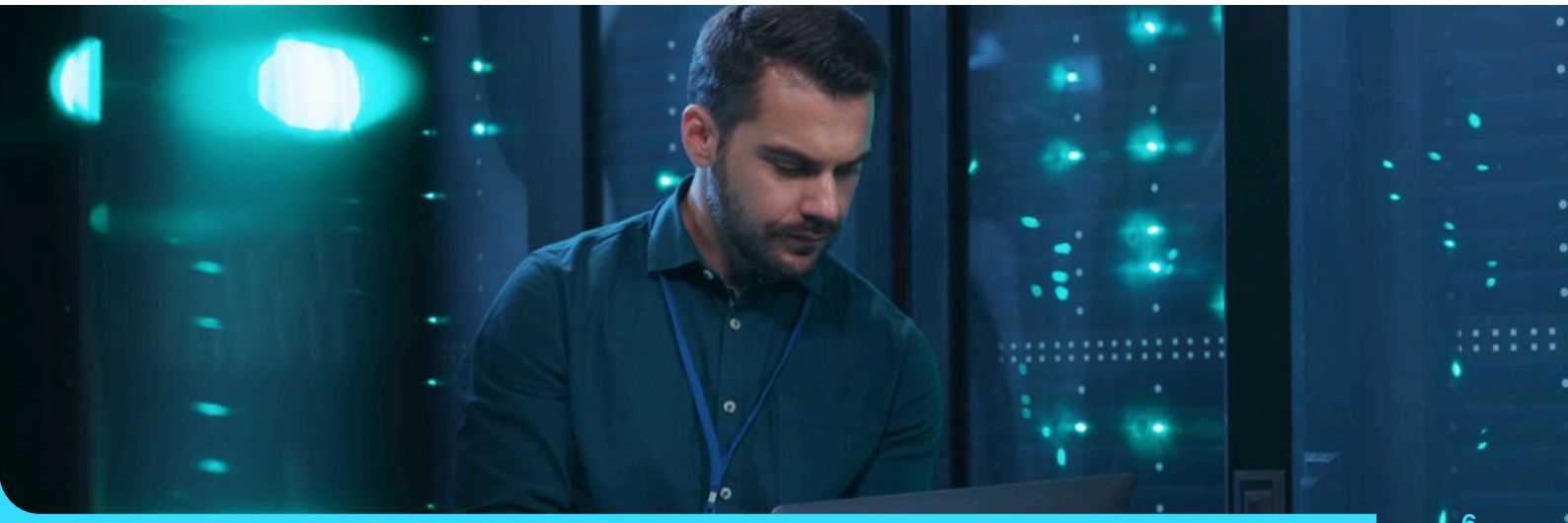
This makes queryable, free DNS data not fully reliable for comprehensive cybersecurity analysis. This problem can only be solved by using global, multi-sensor datasets together with passive DNS data.

Incomplete data

In the case of WHOIS data, a `whois` command sends a basic query to a single default WHOIS server – usually, it's whois.iana.org, which provides a referral server for the appropriate TLD. It might follow up with a query to this referral, but it often doesn't follow through to the registrar's WHOIS server, where most of the detailed domain data can be found.

This results in getting incomplete or possibly outdated WHOIS data. Some registries only store what's called thin WHOIS data, lacking contact details or ownership information, which is only available from the registrar. Ideally, you'll need to query first the registry, and then the registrar to get complete WHOIS information. That makes the throttling problem described above even more acute, as it requires twice the number of the WHOIS requests, as well as some parsing of the results at both stages.

Other data such as NRD feeds is even harder to get if you want to have a full picture. It may require approvals for daily or periodic zone file access requests from registries, and an approval from one registry only gets you access to newly registered domain data for specific TLDs, not all of them. Some TLDs don't offer access to daily zone files at all. Setting up your own DNS sensors to collect this type of information provides limited and incomplete passive DNS perspective as well as risks of mixing up newly observed domains with newly registered ones.



Lack of real-time access

With data such as NRDs, another major challenge is obtaining real-time visibility. Getting data about domains right after registration is necessary for timely emerging threat detection, as many phishing, fraud, and malware campaigns are most active within the first few hours after domain registration. Without real-time NRD access, products lose the opportunity to detect and block these threats before they impact users.

However, even if you manage to get access to daily zone files, they are just that – daily, not real-time. Achieving true real-time visibility into newly registered domains generally requires commercial solutions with privileged registry access, live monitoring of DNS activity, or real-time certificate transparency log tracking — capabilities free sources simply cannot offer.

THE PROS AND CONS OF BUILDING INTERNAL INTERNET INTELLIGENCE FEEDS

After encountering the limitations of free Internet infrastructure data sources, a logical next step is to consider building and maintaining proprietary internal data collection systems.

Indeed, doing so would give you full control over data, its sources, freshness, formats, and integration into the cybersecurity product stack. However, while the benefits are obvious, this approach comes with its own significant challenges.

Parsing and normalizing data

Internet intelligence data is highly unstructured and inconsistent by nature. For instance, different WHOIS servers output information in different formats. This challenge is partially solved by RDAP, which outputs JSON in a standardized format. However, because RDAP supports registry-specific extensions and includes optional fields, field names and structures can still vary slightly across registries. DNS records can also have subtle differences depending on the resolver's location and configuration.

Parsing and normalizing this data into consistent, usable feeds is a non-trivial and resource-heavy task — especially as registry policies, formats, and standards continue to evolve. It may be even harder with the NRD data, which differs even more from source to source.

Complexity and costs

Internal data feeds are often perceived as a cost-effective alternative to commercial feeds. However, in reality, it's often the opposite: setting up reliable Internet intelligence feeds internally requires significant engineering resources, ongoing maintenance, and financial investment.

For example, maintaining a global DNS sensor network, processing high volumes of WHOIS and RDAP queries, or ingesting certificate transparency logs at scale are complex and infrastructure-heavy tasks. Costs include not only the initial setup, but also ongoing expenses related to servers, storage, bandwidth, and, most importantly, the staff needed to process all this data – engineers and analysts.

The table below gives a ballpark estimate of the annual cost of building and maintaining an in-house Internet intelligence program.

COST OF MAINTAINING INTERNAL INTERNET INTELLIGENCE FEEDS	
Factors	**Annual Cost (Range)**
Technology and infrastructure ● Servers, storage, and software for data processing, analysis, and visualization ● Systems for aggregating and correlating data ● Software licenses ● Maintenance	**$150,000 to $500,000**
Personnel (salary and certifications) ● Database & Infrastructure Engineers ● Data Scientists ● AI Engineers ● Cyber Threat Intelligence Analysts	**$1,200,000 to $3,500,000**
TOTAL	**$1,350,000 to $4,000,000**

It shows that developing and maintaining internal data feeds can cost millions of dollars per year. Meanwhile, the annual subscription cost of commercial data feeds ranges from a few thousand dollars (basic feeds) to $100,000–$250,000 (premium feeds). This investment is minimal compared to the millions of dollars required to build and keep an in-house data feed running.

Limited historical data

Another major challenge of building internal feeds is that historical data, which is critical for many cybersecurity use cases, is simply unavailable at the beginning. You can only start collecting data once the systems are operational. This means that for months or even years, the in-house data will lack historical depth necessary for infrastructure mapping, threat actor attribution, or change tracking over time.

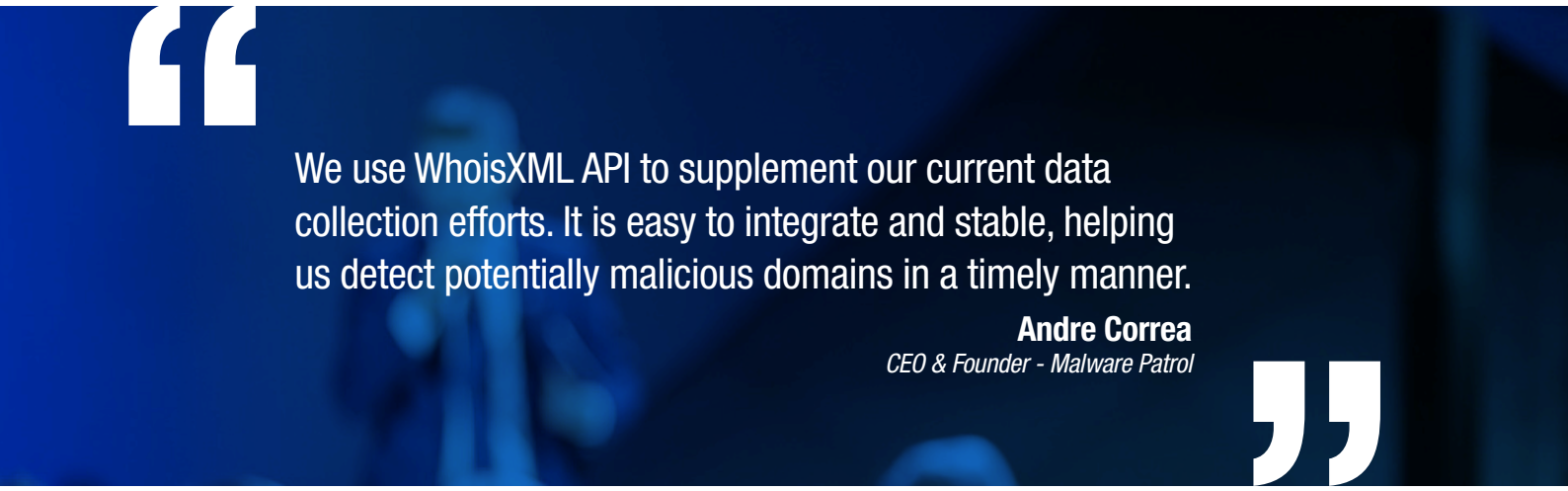
Waiting for years to accumulate sufficient coverage is hardly a viable solution. So, if you need historical data that you don't currently have, commercial data sources remain the only option.

Distraction from building the core product

Finally, but, perhaps, most importantly, developing and maintaining in-house Internet intelligence feeds can easily become a major distraction from the product's primary mission. Product and engineering teams risk focusing on solving data collection and processing challenges, which commercial data providers have already built expertise in, instead of delivering features that actually differentiate their cybersecurity solution in the market.

Internet intelligence collection is, in itself, a full-scale product and operational effort, and splitting focus can slow time-to-market, increase technical debt, and dilute resources.

An internal Internet intelligence feed may still be worth building in some cases, but it's important to understand that internal feeds are not a cost-effective alternative to commercial feeds. Instead, they can be a great solution when there is no commercial product that matches your unique needs.



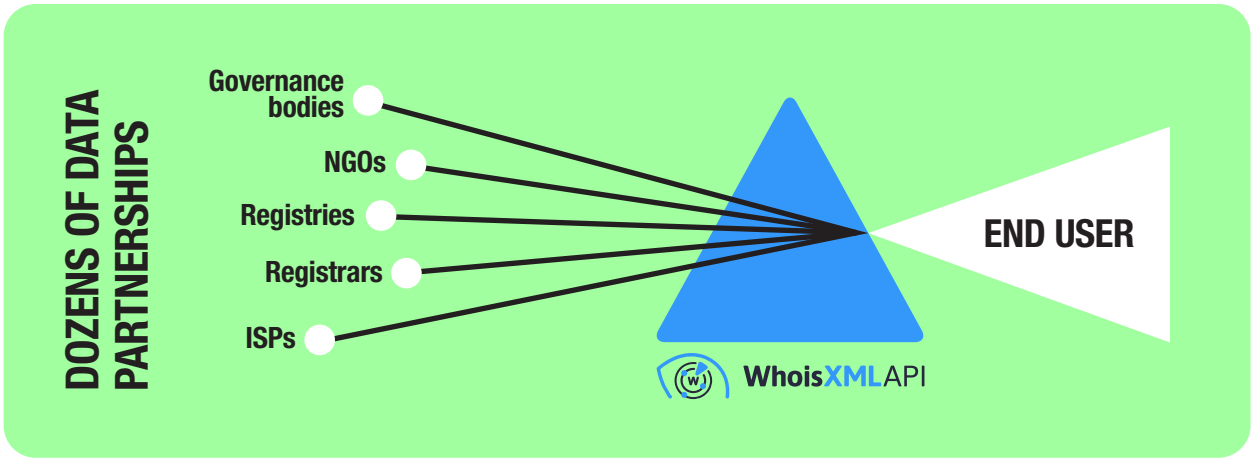
THE BENEFITS OF COMMERCIAL INTERNET INTELLIGENCE SOURCES

While building internal data feeds can provide control and customization, commercial Internet intelligence sources offer substantial advantages that are difficult to match in terms of cost, speed, and operational simplicity.

Access to broader and richer datasets

Commercial providers have the ability to form data exchange and other partnerships with registries, hosting providers, certificate authorities, ISPs, and other data sources at a scale that would be very difficult to replicate internally.

For example, WhoisXML API is a member of Asia and Pacific Top Level Domain Association (APTLD), has partnerships with registrars and actively participates in ICANN policy discussions. That allows it to aggregate data from multiple sources as well as to always be deeply embedded in industry trends and continuously enhance its data products to match the industry developments and challenges.



Enriched data

As certain commercial Internet intelligence providers offer more than a few types of data — WHOIS, DNS, passive DNS, NRD, SSL certificates, IP netblocks, and others — they can combine this data into unified, enriched datasets. This aggregation results in richer, more comprehensive intelligence that significantly improves the accuracy and effectiveness of cybersecurity products.

For example, unlike free NRD feeds, WhoisXML API's NRD feeds contain WHOIS information for each listed domain, making them significantly more usable.

Service level agreements (SLAs)

Another key benefit of commercial sources is the availability of service level agreements, as providers typically guarantee uptime, data freshness, and delivery frequency.

This predictability is critical for cybersecurity products, as it offers something their customers value — reliability. For product managers and engineering teams, it allows planning product capabilities and user experience around reliable data pipelines rather than constantly handling collection issues.

Support and customer service

Commercial data providers offer technical support, assistance with data ingestion, best practice advice, and help resolving integration challenges. In case of technical issues, updates in protocols like RDAP, or changes in registry policies, having a dedicated support channel reduces or avoids downtime and ensures that your cybersecurity product continues to function properly without major disruptions.

Simpler and faster integrations

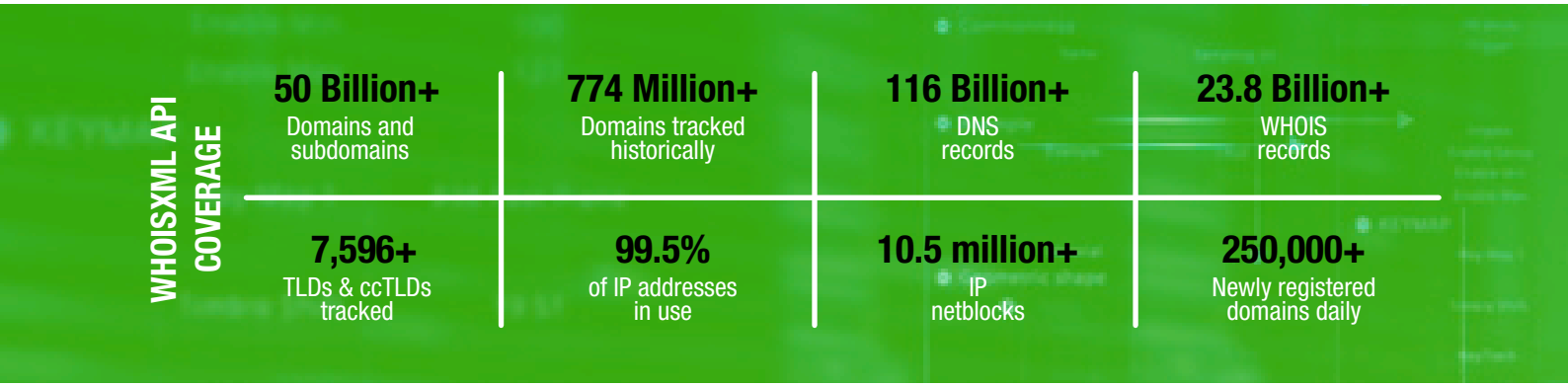
Most commercial feeds are already formatted for easy integration. Whether via APIs, structured data downloads, or database-ready exports, these feeds are designed with customers in mind. They often come with consistent schemas, documentation, and SDKs, which significantly reduces the engineering effort needed to operationalize the data. Instead of building massive parsing and normalization infrastructure, your product teams can focus directly on feature development and threat detection logic.

WHY CHOOSE WHOISXML API INTERNET INTELLIGENCE DATA

WhoisXML API has been providing a comprehensive suite of commercial DNS, domain, and IP data feeds and solutions for over 15 years. These reliable, timely, and SLA-protected data feeds are designed to be valuable and high-quality additions to your current data stacks.

Comprehensive coverage

WhoisXML API's data covers almost the entire Internet, offering insights into almost every domain, subdomain, or IP address. This comprehensive coverage ensures that in almost all cases the product relying on this data would get the context necessary to make correct decisions.



Over 15 years of historical data

Over the course of the company's existence, WhoisXML API has accumulated vast repositories of WHOIS and passive DNS data. That includes over 23.8 billion historical WHOIS records as well as over 116 billion historical DNS records. This historical data provides valuable context for threat analysis.

Diverse delivery models

Different products and use cases require different access to data. WhoisXML API is flexible in this regard and offers access to its data in all possible forms: lookups and dashboards, APIs, databases, and feeds, including monthly, weekly, daily, and live streaming feeds.

Whether you need real-time access, periodic updates, or custom delivery formats, WhoisXML API offers data products that align with your operational requirements.

One provider for all necessary Internet infrastructure data

WhoisXML API offers over 50 different Internet intelligence products, ranging from WHOIS, DNS, and SSL certificates data to threat intelligence feeds, email verification, and brand or domain monitors.

It allows you to choose just one provider for all your Internet intelligence needs, reducing time, complexity, and bureaucracy and allowing you to focus on what matters most.

OEM-only data provider

Unlike many other vendors, WhoisXML API doesn't offer cybersecurity platforms such as threat intelligence platforms, external attack surface management tools, threat exposure management tools, etc.

As a result, WhoisXML API is not competing in the cybersecurity platform market, remaining an OEM data provider focused solely on the one goal of making this data the most comprehensive, consistent, and easy to work with on the market.

Availability of diverse higher-level products for varying needs

Using its comprehensive data repositories and feeds together with advanced machine learning algorithms, WhoisXML API created higher-level data products including predictive threat intelligence feeds that can be used as a first line of defense.

These products benefit from the broad coverage that WhoisXML API has and are easy to integrate into cybersecurity products to enhance their threat detection and prevention capabilities.

FINAL THOUGHTS

Choosing the right Internet intelligence data sources is one of the most impactful decisions in cybersecurity product development. While free data can help during prototyping or initial research, it quickly shows its limitations in terms of completeness, reliability, and performance.

Building internal feeds may offer more control, but doing so requires a long-term investment in infrastructure, specialized engineering talent, and operational resources — often at a cost that far exceeds that of commercial alternatives and, most importantly, diverts the focus from the actual product development.

Commercial Internet intelligence providers offer a compelling value proposition by delivering enriched, real-time, and historical data with strong SLAs, built-in integrations, and dedicated support. For most cybersecurity products, this means faster time to market, improved threat coverage, and more resources focused on innovation rather than infrastructure.

As a specialized OEM data provider with unmatched global coverage and historical data, WhoisXML API offers a broad variety of Internet intelligence data solutions that form a strong foundation for cybersecurity products currently leading the market.

ABOUT US

WhoisXML API provides well-parsed, normalized, and comprehensive WHOIS, IP, and DNS intelligence. For more than 15 years now, we have gathered and aggregated 23.8+ billion historical WHOIS records, 50+ billion hostnames, 116+ billion DNS records, 10.4+ million IP netblocks, and 99.5% of active IPv4 and IPv6 addresses in use.

WhoisXML API has more than 52,000 satisfied customers from various sectors and industries, such as cybersecurity, marketing, law enforcement, e-commerce, financial services, and more. For several years now, it has been recognized as an Inc. 5000 honoree and one of the Financial Times's Top Fastest-Growing Companies.

Visit whoisxmlapi.com or contact us for more information about our products and capabilities.

WhoisXMLAPI
The Who Behind Domain, IP & Cyber Threat Intelligence