ABOUT ONCRAWL

We are a team of 50 passionate explorers based in Bordeaux & Montreal and backed by leading investors. We collect 25M+ URLs & 1B+ log lines every day for 1000+ clients ranging from SMBs to Fortune500 companies.

It is the shared interest for search engines and the exploitation of Big Data that led François Goube and Tanguy Moal to join forces to create Cogniteev, the company behind the data platform Oncrawl. With over 20 years experience in search and semantics, François Goube, serial entrepreneur, and Tanguy Moal, NLP expert and Data geek wanted to build the most advanced research technologies accessible to the greatest number of people. And it was already proven from their previous experiences that they were able to build products used by millions of people (Exalead, JobiJoba...).

In 2013, Oncrawl was born. Built from an industrial consortium with Cdiscount – the leading French e-commerce website – the team behind Oncrawl managed to develop a SEO crawler able to analyze millions of pages – in a short period of time – and at a competitive cost. They then understood that a large number of SMBs but also larger companies needed that technology previously held by Fortune500 companies. That’s why they decided to democratize such a solution by officially releasing Oncrawl in 2015.

Oncrawl provides a scientific approach to search engine optimization by delivering reliable and accurate data that can be used to monitor, understand and optimize a website’s performance on search engines results.
Our mission

Oncrawl is a **technical SEO platform powered by data science** that helps search marketers make smarter SEO decisions. We help e-commerce websites, classifieds and online publishers to improve their organic traffic, rankings, and online revenues by **opening Google’s black box**. Driven by a strong tech spirit, Oncrawl gives search marketers easy access to the data they need by providing actionable dashboards and reports to support the entire search engine optimization process. In just 3 years, Oncrawl has won nearly 15 search awards, making it the most awarded SEO platform on the market today.
Our crew

The Oncrawl team is composed of 50 passionate and talented people based in Bordeaux and Montreal and we are always looking to hire new talents throughout the year, one sign of Oncrawl’s extensive growth.

François GOUBE
CEO and founder

“We have come an extraordinarily long way in a remarkably short time. Whether it is about technology, support or services, Oncrawl has become much more than a technical SEO platform and always goes the extra mile to impress its users. Only this year, we have won nine search awards, including the Best Global SEO Software Suite, released exclusive data science features, onboarded enterprise customers and we have built strong relationships with the international SEO community. We will keep working hard to improve Oncrawl and to reach our goal: become the leading technical SEO platform on the market.”
Tanguy MOAL  
CTO and founder

“Oncrawl is one of the most complete and strongest SEO solutions available on the market. Our solution is auto-scalable and offers an unlimited cloud-based storage capacity. The technologies used by Oncrawl allow our users to crawl at an unlimited speed.

Plus, we give great importance in building a secure data system which is completely GDPR compliant. Because research and development is at the very heart of our DNA, we continue to expand our engineering team in order to always offer the latest advances in terms of massive data analysis.”

Vincent Terrasi  
Product Director

“Oncrawl is the only SEO platform which integrates data science technology to help companies predict their online performance. Using a Machine Learning model allows entrepreneurs to find the important variables that impact their website and to efficiently predict their most important KPIs such as visits, clicks or revenues. Having access to this exclusive and cutting-edge technology is a clear competitive advantage and our team continues to work hard to provide the most innovative features.”
Our values

Our team is made up of 50 passionate Oncrawlers. We offer a constellation of benefits and perks to help our team strike just the right balance between work time and personal time.

- **Innovation & Excellence**
  With performance, scalability and reliability in mind, we implemented our own web scale SEO crawler with strong natural language understanding and user behaviour modelization features.

  Our solution leverages the power of cloud computing to bring the required scalability (both up and down) to fit any increase in demand and reduce costs. We’ve worked really hard to merge data science and SEO and to offer unique features tackling challenging SEO issues. That’s why we’ve been the first to release a near duplicate content detector or an internal page rank, measuring internal popularity flows.

- **Authenticity & Kindness**
  We describe ourselves as the good guys of the industry, and bringing the best customer service possible is our priority mission. We believe that our customers are our first ambassadors and that delivering the best relationship is the key to building long-term relationships. People at Oncrawl are smart, friendly and care about what they do.

  We also work really hard to provide actionable resources that can help the SEO community understand the value of technical SEO and data science and what they can do to achieve their goals. That’s also why we open-sourced a few of our developments on Github like our ELK log analyzer to offer easy and free access to strategic features.

- **Competence & Ambition**
  Oncrawl has been built by Big Data pioneers and search engine experts who began working in distributed software and web scale algorithms in the early 2000s. Oncrawl is an SEO solution that is developed by real SEOs who understand the challenges the industry is facing. We plan to become the leading cloud-based technical SEO on the market and we are convinced we have the right assets to do it.
Our expertise

The company relies on five major areas of expertise, completely mastered in-house:

• **Crawling**: for nearly 20 years, the founders have been involved in or have developed search engines (Exalead and JobiJoba in particular). The company now has the ability to browse and index the web in near real time.

• **Big Data processing**: Oncrawl develops not only processes for capturing and homogenizing data from the web, but also machine learning processes that lead to future self-learning systems, eventually enabling data mining operations to be carried out on a web scale. Current analyses are based on algorithms that support huge datasets and massive sites, allowing our solution to fully support giant projects as a matter of course.

• **Semantic analysis**: as specialists in NLP (Natural Language Processing), Oncrawl’s teams have developed an entity detector to «capture» data from the web on all types of entities, both known (Freebase.org, Open Data...) and unknown (machine learning). An automatic classifier allows the homogenization of unstructured data.

• **Distributed architecture**: given the volume of data and content processed, the company’s applications require the implementation of an n-tier architecture. The company therefore has perfect control over the scale-up of its services, and deploys all its developments on cloud-based hosting.

• **Data Science**: with the release of Oncrawl Labs, the platform has proven its expertise in data science. Oncrawl is recognized in the industry for connecting the dots between technical SEO and data science. The Labs is entirely based on Data Science and Machine Learning and allows companies to predict their online performance and automate a part of their strategy with pre-made algorithms developed with Python and R.
Reaching the stars since 2013

2013
- Birth of Cogniteev, the company and data platform behind Oncrawl Pilot project with Cdiscount’s industrial consortium

2014
- Partnership with two research labs
- $650K Seed Round investment closed

2015
- 1.5M€ spent in R&D to develop cutting-edge features
- Release of Oncrawl’s public version
- 100+ clients

2016
- ELK Open Source log analyser available
- Oncrawl Cloud-based log analyser available

2017
- First Best Search Software Tool at the European Search Awards
- AT Internet & Majestic partnerships
- Oncrawl 3.0 new version released
- Offices opening in Montreal, Canada
- 600+ clients

2018
- Search Console, Splunk, Adobe Analytics integrations
- 3 prizes won at the European Search Awards, SEMY awards and Grand Prix du Search
- $4M raised to accelerate our international expansion

2019
- Winner for the 3rd year in a row of Best Search Software Tool at the European Search Awards
- Winner at the US & MENA Search Awards
- 1000+ clients
- An iPad filled with SEO tips was sent into space
- Move to bigger offices

2020
- Release of Oncrawl Labs, an R&D platform exploring the intersection of technical SEO, data science and machine learning.
- Winner at the Global, Canadian, UK, US and MENA Search Awards
- Release of our new plugins with Dataiku and Google Data Studio
- First charity campaign with Plan International

2021
- Most awarded SEO platform on the planet

2022
- ...More to come!
Our awards

- 2014 TheNextWeb Award
- 2016 Digital In-Pulse Award
- 2017 Best Search Software Tool at the European Search Awards
- 2018 Next’ Step Award
- 2018 Jury Special Award at the SEMY Awards
- 2018 Best SEO Software Suite at the European Search Awards
- 2018 Best Optimisation Platform at the Grand Prix du Search
- 2019 Best Search Software Tool at the US Search Awards
- 2019 Best Search Software Suite at the MENA Search Awards
- 2019 Best Search Software Tool at the European Search Awards
- 2020: Best SEO Software Suite Canadian Search Awards
- 2020: Best Global SEO Software Suite at the Global Search Awards
- 2020: Best SEO Software Suite at the US Search Awards
- 2020: Best Innovation Software at the US Search Awards
- 2020: Best Search Software Tool, Best Software Innovation and the silver award for Best Best SEO Software Suite at the MENA Search Awards
- 2021: Best Software Innovation at the APAC Search Awards
- 2021: Best SEO Software Suite at the APAC Search Awards
OUR PRODUCTS & SERVICES

Oncrawl is a technical SEO and data platform that offers a broad range of features and more than 600 metrics relying on dozens of ranking factors to help search marketers to better optimize their site with regard to Google’s guidelines.

Our products

Oncrawl SEO Crawler

Oncrawl SEO Crawler enables users to analyze their websites in the same way as Google’s bot does, no matter how large the website is. We provide our customers with comprehensive reports based on reliable data which evaluate the quality of content, HTML code and architecture as well as their website’s performance. These reports are entirely customizable with desired metrics and exportable.
Oncrawl Log Analyzer

Oncrawl Log Analyzer helps search marketers to understand how search engines and visitors behave on every part of their website and if strategic pages are being visited. We provide real-time, scalable and actionable dashboards to anticipate Google Search Console alerts.

Oncrawl’s log file analyzer is GDPR compliant and respects data privacy: users’ IP addresses are not stored within the Oncrawl application, and log data is transferred to Oncrawl using a secure and private ftp account.
Oncrawl Data³

Oncrawl provides seamless connections to additional datasets to enrich its user’s SEO analysis. We help our users understand the positive and negative influence of ranking factors on indexability and SEO traffic.

Oncrawl Data³ enables users to analyze the effectiveness of an SEO strategy by combining crawl data with backlinks, SEO traffic, rankings, and custom datasets from a CRM, monitoring solution, or any other source.

- **Oncrawl SEO Impact**: Lets the user optimize crawl budget and ROI on strategic pages based on analysis from their crawl data and log files.
- **Oncrawl Rankings**: Lets the user analyze the influence of ranking factors on positions and impressions based on analysis from their crawl and search console data.
- **Oncrawl Analytics**: Lets the user drive SEO, SMO traffic and users to top-priority pages based on analysis from their crawl data and Google Analytics, AT Internet or Adobe Analytics data.
- **Oncrawl Backlinks**: Lets the user analyze the influence of their backlink distribution on crawl frequency based on analysis from their crawl data and Majestic data.
Oncrawl Platform

Oncrawl Platform manages and processes any dataset and customizes analyses to build unparalleled SEO reports based on Big Data intelligence. But it is also fine-tuned to any SEO strategy and to a website’s profile and size, drawing from real data, whether it’s data exported from the latest technical tool or data users collect by hand. Everything is exportable and explorable.
Oncrawl Connectors

Oncrawl allows its customers to combine their crawl data with trusted search marketing and data science solutions for deeper SEO insights.

- **Google Analytics:** understand the impact of SEO on your traffic and on user behavior
- **Majestic:** manage backlinks activity and their impact on SEO performance
- **AT Internet:** control your tagging plan and analyze organic traffic
- **Adobe Analytics:** understand how on-page and technical SEO affect organic traffic performances from search engines
- **Google Search Console:** monitor indexing rate and influence of SEO factors on positions
- **Google Data Studio:** Enrich your crawl data with Business Intelligence and Data Visualization to go further in SEO audits or business analysis.

Oncrawl has also partner with leading search marketing and data platform to give access to its SEO data

- **Dataiku:** get access to SEO data in Dataiku, where it is ready for use in data science and machine learning projects.
- **AT Internet:** Combine your AT Internet data and crawl data to understand how analytics performances impact your website on search engines.
Oncrawl Labs

Oncrawl Labs is an R&D platform exploring the intersection of technical SEO, data science and machine learning. Developed on Google Colab and using Python and R languages, Oncrawl Labs offers a portfolio of algorithms as solutions to address strategic SEO issues, gives access to features not yet available on the SEO market and feeds Oncrawl’s product roadmap. Here are the four algorithms available for now:

- **Real-Time Indexing**: automatically detect new URLs and submit them to search engines for indexing.
- **Long-Tail Prediction**: train time-series prediction models to predict future long tail trends.
- **Text Generation**: test transformers with your data and generate new qualitative texts for SEO purposes

---

**Real-Time Indexing**

**Context**

How to automatically detect new URLs based on the last two crawls and submit them to search engines for indexing.

Getting new pages indexed is a challenge for SEOs in industries with frequently evolving sites, particularly e-commerce, classifieds, and online publishers, where the rapid searchability of new pages directly affects the business.

The major search engines provide means of manually submitting pages for indexing. However, in the use cases described above, this can require SEO teams to maintain extensive daily lists of pages created—sometimes automatically—by production and content teams. Obtaining a complete list can be difficult.

To submit URLs to search engines, we will use the following methods, which you should be familiar with before beginning:

- How to make your sitemap available to Google? [https://support.google.com/webmasters/answer/1836667?hl=en](https://support.google.com/webmasters/answer/1836667?hl=en)
**Oncrawl BI**

Oncrawl’s direct compatibility with the most common BI solutions allows you to recognize the strategic capabilities of SEO and search performance. By connecting to BI solutions and data visualization software, you not only improve your reporting and dashboards for SEO, but also show how SEO and data from other departments interact.

**Our Customer Success Services**

We’ve created a support team made up entirely of SEOs and engineers with at least 5 years of experience, which means that our support team has strong knowledge of the entire technical SEO chain. They come from different backgrounds – such as in-house or agencies – but they have worked on very large websites with challenging SEO issues.

This enables them to provide accurate, effective and advanced training and support, no matter the complexity of the SEO issues that our users face. We provide real support to our customers and offer plenty of opportunities to grow as a search marketing professional, including a customized kickoff, the Rocket Program training, a workshop to master our solutions, an up-to-date knowledge base and many free webinars.
OUR TECHNOLOGY

Oncrawl uses reliable, secure and auto-scalable technology to meet our users’ SEO needs.

Auto-scalability and cloud-base storage

Oncrawl has an unlimited storage capacity thanks to the most powerful cloud in the world: the Google cloud. Our solution also has access to many data centers across the world to ensure robust and reliable storage.

Data hosting in the Google cloud allows us to adjust our storage according to the volume of data and the number of customers using Oncrawl.

Our solution is auto-scalable: our clusters grow depending on their activity and the allocated resources vary according to the data handled in real time. In other words, Oncrawl’s design allows it to adapt to our users’ needs for storage limits and processing power.

Unlimited crawling capabilities

The technology used by Oncrawl allows us to crawl at an unlimited speed. The only limit is the efficiency of our clients’ infrastructure. Oncrawl is able to adapt its capacity to the speed desired by our customers and adopt a “safe” mode. Currently, we collect up to 25 millions of URLs per day and per website, or approximately 1 billion web pages and 150 billions of links per month.

Secure data system and GDPR compliance

We guarantee comprehensive data security for our customers. Oncrawl is completely GDPR and CCPA compliant and meets all data confidentiality norms. For our European clients, we can ensure data storage on the European continent. We have set up a policy for restricting internal access to data. Every log file that we collect is stored in our isolated and secure FTP space. Our service is protected by the HTTPS protocol and by isolated software components. It also profits from extra protection from a software stack that we maintain up to date.
**Real-time crawl monitoring**

With Oncrawl, our users can check the progress of their crawls and the analysis of their logs in real time. In order to facilitate their work, we have a notification system which keeps them informed of the progress of their SEO analysis. Our customers can check on the status of their data at any moment.

**API Access**

Oncrawl provides a SaaS application which offers a complete analysis of our customer’s data. We know that some of our users would like to build their own dashboards with raw data from Oncrawl and others sources. That is why we offer complete access to our API. Our solution gives total freedom to users who are interested in integrating Oncrawl’s data into their own ecosystem. Thanks to the API, they can go even further with the Oncrawl data!
They’re talking about us

Don’t take our word for it: discover what our customers are saying about us.

“As a team lead focusing on the growth of our multinational platforms, regular reporting/monitoring and automated risk management of product changes’ effect on our organic search performance is very significant and challenging. Oncrawl is becoming an essential part of my team day by day with its capability and accuracy.”

Murat Yatağan, VP Growth at Brainly

“As a technical SEO who does SEO audits for a living, I have to rely on crawlers and SEO tools to gather the data I need to do my job. There are several crawlers out there that I could use, but I prefer oncrawl. When it comes to being able to combine data sources and crawl large sites, I have not found a better tool. I like the robustness and the flexibility that oncrawl gives me.”

Bill Hartzer, SEO consultant
“With Oncrawl it is very easy to bridge the gap between digital marketers and their clients and form long-term relationships.”

Omi Sido, Senior Technical SEO at Canon Europe

“We chose to invest in Oncrawl because we wanted to get ahead of some of the issues that could disrupt Google’s ability to crawl our site efficiently. As a large publication, we need to ensure that Google can crawl, render, and index our content just as quickly as we publish.”

Dayne Richards, SEO Analyst at Forbes

We’ve been featured in

SEJ SearchEngine Journal
Search Engine Watch
STRATÉGiES
JDN Journal du Net
Search Engine Land
The Drum
STATE OF DIGITAL
insideBIGDATA
What is SEO?

The acronym SEO stands for “Search Engine Optimization”. This discipline consists in ranking a website, a web page or an app in the first positions of search engine results. A lot of requirements can have an impact on the website indexing, which is why data analysis is essential for SEOs.

Oncrawl offers an actionable platform which allows SEO experts to identify the positive and negative factors behind the indexing of their websites and to set up necessary improvements. For example, the owners of websites have to pay attention to their content quality, to the architecture of their site, as well as to their internal linking because these different factors directly impact their site’s SEO performance. By studying these and other factors, they can conclude which actions they should carry out to boost the organic (as opposed to paid) visibility of their website: a true advantage for the sites which need to generate a lot of traffic to fulfill their business goals.

What is technical SEO?

The words “technical SEO” refer to the actions implemented to optimize your website and server in order to help search engine bot to crawl and index your site more effectively. Technical SEO gathers many factors such as page elements, HTTP header responses, XML sitemaps, redirects, meta data, etc. It can directly or indirectly impact your rankings. This is a large field but in a nutshell, we can say that technical SEO has 3 objectives:

• Make sure that search engines bots can easily access and index your website
• Give the right signals to the search engine crawlers so they can understand the structure of your website
• Give search engine algorithms reasons to trust your website and rank it higher
What is a crawler?

The term “crawl” refers to the action of the bots sent by search engines to analyze and index a website. From an initial start page or pages, the bots will follow all the links on that page to find new pages on the website. This process repeats on the newly discovered pages until the bots can’t find any new links. At the end of this crawl, the bots will decide if they will index the pages, that is, save them as potential answers to searches. For each search query, the site’s indexed pages may or may not appear in the search results. The site’s position in the results is determined based on its relevance; it is judged by its authority, content, architecture and over 200 other factors!

What is a log file?

A log file corresponds to the saved information on the server each time that there is a new visit on one of the pages of a website. This helps you identify organic traffic and monitor what Googlebot is doing on your website in real time. In SEO, these files are essential because they can validate the return on improvements that have been implemented and identify regressions early on.
CONTACT

Need to get in touch with our PR team to know more about our company or discuss media coverage?

Gaëlle Judes
Head of Communications & Events
gaelle@oncrawl.com