ewdci

Ethical Web Data Collection Initiative (EWDCI) Principles 1.0

EWDCI is a consortium of web data collectors focused on strengthening public trust, promoting ethical guidelines, and helping businesses make informed data collection choices.

The Introduction

Web Data Collectors are companies who gather available data from around the internet, enabling consumer confidence, commercial innovation, and community safety. By web data collection, we mean data that is aggregated and can come from a variety of sources, including the open web and mobile applications.

The Ethical Web Data Collection Initiative (EWDCI) is an industry-led and member-driven consortium of web data collection business leaders focused on strengthening public trust, promoting ethical guidelines, and helping businesses make informed data collection choices.

Enabling Consumer Confidence, Commercial Innovation, and Community Safety.

The collection of web data enables business innovation and competition, provides consumer benefits – including serving accurate information and comprehensive pricing – and facilitates thorough data analysis for academic and public safety purposes. Some use cases of note include:
- E-commerce price intelligence, which includes up-to-the-minute comparative information on same-item pricing across the internet.
- Jobs and real estate listing aggregation.
- Social listening and brand monitoring.
- Alternative data for investment decision making.
- Businesses performing a thorough review of consumer trends and competitor companies worldwide to best position themselves for success.
- Companies scanning internet data for illicit uses of their intellectual property.
- Businesses confirming validity of marketing spend via ad verification.
- Data aggregation companies supplying pro-bono scraped data to academic researchers, providing them vast data stores for essential research.

How We Foster Trust

The Ethical Web Data Collection Initiative brings together first-hand insights and focuses on advocacy, education, and awareness.

We seek to address concerns, and proactively solve associated challenges by:
- Defining data collection best practices
- Designing and reinforcing industry standards
- Providing accurate information to government officials and policy makers
- Promoting appropriate industry-led regulations
- Informing the wider technology industry
- Advocating for responsible collection and use of personal data

To ably provide these benefits to consumers, companies, and the public at large, data collectors need to follow ethical practices. Such dedication will serve to further the understanding of data collection practice and process, and further the adoption of this technology. Key to this is a safe and community-minded application of collection, and therefore a positive reputation for the practice and its practitioners. The EWDCI gathers leading voices within the industry to determine and advocate for such integrity in aggregation.

Bringing Together Web Data Collectors to Build Trust.

The EWDCI is an initiative of the i2Coalition, the leading voice for web hosting companies, data centers, domain registrars and registries, cloud infrastructure providers, managed services providers, and related tech.

The EWDCI was formed in 2022 when i2Coalition brought together five companies in the web data collection space to participate in a new initiative aimed at fostering cooperation in the industry. The EWDCI was established by five founding member companies and their associated brands: Coresignal, Oxylabs, Smartproxy, Rayobyte, and Zyte. The EWDCI is actively seeking other members of this field dedicated to best practices to join and help shape the initiative's early efforts.

The EWDCI is dedicated to serving as the voice of the industry, collaboratively strengthening public trust in the practice of data collection, promoting ethical guidelines, and helping businesses make informed data collection choices.

The Principles

01 Legality

Web Data Collectors will only scrape and use tools to scrape data within the confines of legislation and regulation from relevant jurisdictions, and will strive to keep up to date with judicial cases that may affect the legal boundaries of the work.

Relevant Laws - In order to maintain awareness of legal pitfalls, web data collectors should keep abreast of copyright law, anti-hacking laws, contract law, data protection laws, and any other relevant local or regional laws that may be tangential to their business, including but not limited to the California Consumer Protection Act and the European Union's General Data Protection Regulation.

Privacy – web data collectors shall strive to comply with privacy laws and respect and protect the privacy of internet users whose data is being collected

Proxies – Web data collectors will maintain the legality of the proxies they use in the process of collecting data from the internet

02 Ethics

Companies engaged in collecting web data will follow a strict ethical framework, including ethical principles related to four main pillars: websites, customers, proxies and data.

Websites – Web data collectors will set rate limits and scraping targets in a way that is beneficial to both the collector and the target.

Customers – Web data collectors will strive to conduct Know Your Customer checks on customers using residential IPs and other high risk data collection tools to ensure that the data will be used ethically and services provided to legitimate companies.

Proxies – Web data collectors will strive to use only ethically sourced proxies, through informed consent processes, and in accordance with the data protection laws in the relevant jurisdictions.

Data – Web data collectors will collect internet and web app data in an ethical manner, being mindful of the potential misuse of such data.

03 Ecosystem Engagement

Web Data Collectors are in a symbiotic relationship with the free and open internet ecosystem and strive to engage it collaboratively in an open and communicative manner.

Supporting data accessibility – Web data collectors believe in the principle that data should be open and accessible and will strive to do so as an industry, in order to nurture businesses and entrepreneurs who build on top of the open internet.

Healthy competition – Web data collectors believe in collaborative engagement both within and outside of their industry, establishing industry standards and abiding by the commitment to having a seat at the table in matters of the health and prosperity of the internet.

04 Social Responsibility

Web data should never be used to the detriment of society, and members pledge to support and collaborate with civil society and governmental organizations for societal benefit.

Societal benefit – Web data collectors define societal benefit to include use by educational institutions as well as non-profits, and use for environmental causes.

Donate data – Web data collectors pledge to donate, and if applicable analyze, data when appropriate, to its societal benefit collaborators.
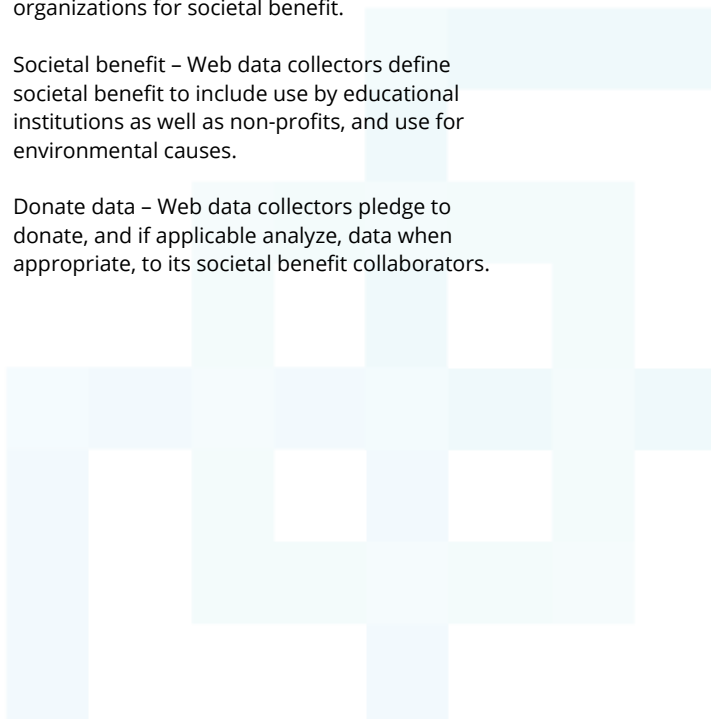
Web Data Aggregation

The public deserves digital peace of mind, *and "scraping" doesn't have to be a dirty word when it is done responsibly, but responsibility needs defining.*

The web data aggregation ("web scraping") industry is still young, and still growing. As with any industry in its early stages, those working in this field have a one-time opportunity to shape how it is developed and perceived.

This industry will not stand to allow bad examples to define it, and so we're organizing and taking action.

The Ethical Web Data Collection Initiative (EWDCI) is an international, industry-led consortium of web data collectors focused on strengthening public trust, promoting ethical guidelines, and helping businesses make informed data aggregation choices.

The EWDCI is dedicated to defining positive and beneficial uses of the important abilities and potential of web data collection and aggregation at scale.

What we do:

- Advocate for responsible web data collection and use of personal data
- Educate and guide the industry on the use of ethical resources and tools used in web data collection
- Foster consumer confidence in data collection through transparency and accountability
- Enable commercial innovation
- Promote online safety

Our goal is to prevent harmful legislation from passage worldwide, but also, potentially to seek inclusion in federal laws. In addition, we are building a framework to establish an open, participatory process around the development of legal and ethical web scraping provider principles, and then hold industry leaders accountable for those ethics. **This work is ongoing and needs your insight!**

Web scraping can be an ethical practice that invites progress for the free and open Internet, but it requires leaders to step up and make it so. Join the Ethical Web Data Collection Initiative (EWDCI): https://ethicalwebdata.com/

About EWDCI + i2Coalition

The Ethical Web Data Collection Initiative (EWDCI) seeks to foster cooperation in the web data collection and aggregation industry and leverage collective first-hand knowledge and insights to advocate for beneficial technical standards and business best practices regarding the aggregation of data. The EWDCI is dedicated to serving as the voice of the industry, collaboratively strengthening public trust in the practice of web data collection and aggregation, promoting ethical guidelines, and helping businesses make informed data aggregation choices.

The Internet Infrastructure Coalition (i2Coalition, i2C) is the leading voice for web hosting companies, data centers, domain registrars and registries, cloud infrastructure providers, managed services providers, and related tech. The i2C works with Internet infrastructure providers to advocate for sensible policies, design and reinforce best practices, help create industry standards, and build awareness of how the Internet works. The i2Coalition also spearheaded the creation of the VPN Trust Initiative, which determined and promoted best practices for that vital industry.

Web Scraping Fact Sheet

ewdci

What is "web scraping"?
In essence, web data aggregation ("web scraping") is simply going to a web page, extracting data from it, and sorting that data to make it understandable. When done at scale, the results can be powerful. Whenever you search for something on Google, what you see is the result of web scraping. So you see that there are some very compelling positive use cases.

How does it work?
If you've ever copied and pasted something from a website, you've operated as a web scraper—but of course we're talking about scale. Modern web data aggregation is done with bots and database software to gather large amounts of data, parse that data, and present it in a way that's understandable to humans. You can find open-source data aggregation tools out there; as well as proprietary tools and companies that specialize in web scraping.

Wait, What's a Bot?
A bot is an autonomous program that performs (often repetitive) tasks on the Internet much faster than a human could. There are good bots and bad bots, based upon how they're configured and how they're deployed. There are a lot of bots out there—over 40% of Internet traffic is made up of bot activity. Examples of good bot behavior: Web crawlers, customer service chatbots. Examples of bad bot behavior: Spambots, DDoS attacks, click fraud automation.

What can be done with aggregated web data?

- Search engine functionality
- Gauging customer sentiment
- E-commerce competitor analysis
- Hotel and flight price comparison
- Market research
- Academic research
- Real estate listings
- Weather data monitoring
- Search engine optimization (SEO)
- Website change detection
- Online reputation management
- Data visualization

Is web scraping legal?
Yes—in fact, the modern Internet wouldn't work very well without it! Different jurisdictions have laws around which types of data can be aggregated, how it can be aggregated, and how that data can be used. For example, The European Union's General Data Protection Regulation (GDPR) is a widely known example of this, but there are many more—and new laws are being written as we speak.

Web Scraping Does Not Equal Hacking
Web scraping involves extracting data from a website and storing it in a structured format. Hacking involves unauthorized access or manipulation of a computer or network. If web scraping is the equivalent of street photography, hacking would be like setting up a camera in someone's house.

Why does web data aggregation often have a bad reputation?
Web data aggregation can be done maliciously. For example, web scraping is a core component to DDoS attacks. Also, the collection of personal data at scale often runs afoul of both local laws and the bounds of accepted business behavior. However, this bad reputation comes from a few bad actors and most web scrapers are conducting ethical web data aggregation for one of the reasons listed above.

What is the good that comes out of it?
When collected and sifted in certain ways, a dizzying multitude of anecdotes becomes actionable data. As companies and individuals, we can accomplish much more when we're able to make sense of the vast amounts of disparate bits of information out there. However, like any technique, web scraping can be misused. That's why we're forming this group: to prevent bad actors from sullying an industry that can accomplish a whole lot of good.

The Public Deserves Digital Peace of Mind
The Ethical Web Data Collection Initiative (EWDCI) is an international, industry-led consortium of web data collectors focused on strengthening public trust, promoting ethical guidelines, and helping businesses make informed data aggregation choices.

The EWDCI is dedicated to defining positive and beneficial uses of the important abilities and potential of web data collection and aggregation at scale.

What we do:
- Advocate for responsible web data collection and use of personal data
- Educate and guide the industry on the use of ethical resources and tools used in web data collection
- Foster consumer confidence in data collection through transparency and accountability
- Enable commercial innovation
- Promote online safety

Our goal is to prevent harmful legislation from passage worldwide, but also, potentially to seek inclusion in federal laws. In addition, we are building a framework to establish an open, participatory process around the development of legal and ethical web scraper provider principles. "Scraping" doesn't have to be a dirty word.