

# Taking Complete Command of Web Data

*How businesses are moving from risky home grown “scraping” solutions into secure, comprehensive Web Data Integration (WDI)*

**Web data** is the boundless unstructured data living on websites, web portals and SaaS applications

---

Some **API access processes can take weeks, months, or longer** to approve and set-up

## The current state of web data

Web data is the boundless unstructured data living on websites, web portals and SaaS applications—simply put, it’s the data in your web browser. For years, web data has been something you can see, but not really “touch” or incorporate into your business systems without accessing each website via an API or building homegrown scraping solutions. But that’s about to change.

## Finding the fuel to move forward

Often, the data a business needs to fuel success can be accessed by APIs, but not always, as there are countless non-API sources of information on websites, web portals, and even some SaaS applications.

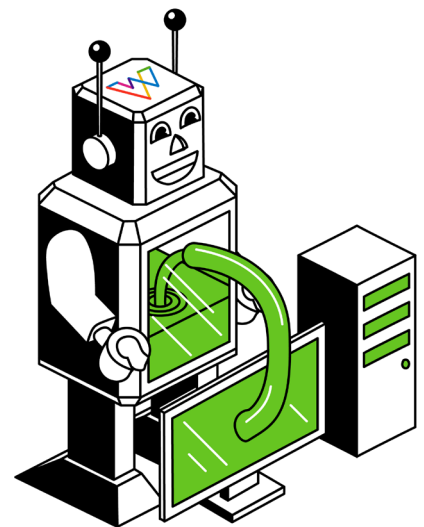
Take for example many Mobile Telecom Companies who offer APIs to customers to help manage expenses, but leave out critical pieces of data, like current international usage, which is invaluable for avoiding costly roaming charges. This usage data may not be in the API, but is instead sitting as unstructured web data, trapped in a customer web portal...ripe for consumption if it could only be easily accessed at scale.

## Moving past relying (solely) on APIs

While APIs are a standard means of data transfer, that doesn’t necessarily make them the most expedient or ideal. Some API access processes can take weeks, months, or longer to approve and set-up, and these delays raise risks of failed deployments, missed revenue opportunities, and poor customer service.

But assume API setup goes smoothly: what if your API vendor sunsets support, or decides to make significant changes that make it more difficult to work with their API?

The ability to independently access web data moves from being a “nice to have” to a must, as does the need for a modern Web Data Integration (WDI) tool that can extract web data across multiple websites with the same scalability, reliability and structure of an API.



According to 2019 research from Opimas, spending on web data extraction was \$2.5B in 2017 and predicted to reach nearly \$7B by 2020, with the majority of spend on internal homegrown solutions. But while the overall web data extraction spend is trending up, Opimas expects a significant transition from internal spending (with expected growth at 30%) to external spending (with expected growth at a whopping 70%).<sup>(1)</sup>

So why the move toward more external solutions?

Spending on web data extraction was **\$2.5B in 2017** and predicted to reach nearly **\$7B by 2020**  
— Opimas Research

In their current state, internal solutions for web data extraction are often unstable and send valuable software development resources down a proverbial rabbit hole of custom coding and neverending maintenance in an attempt to manage: data extraction templates for every data source, IP blacklisting, CAPTCHA, 2 factor authentication, pacing, data access across multiple pages / websites or data behind logins, file downloads and scheduling — to name just a few.

*Web Data Integration,  
Leveraging the Ultimate  
Dataset*

Signs are pointing toward use of modern Web Data Integration (WDI) solutions that quickly transform any unstructured web data into structured APIs — with no coding required. Specifically, faster, more reliable, more scalable, and more affordable alternatives to legacy homegrown and outsourced web data extraction solutions.

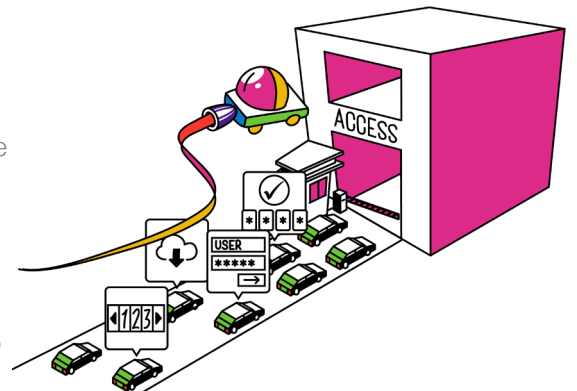
## How modern Web Data Integration (WDI) differs from “scraping”

Modern Web Data Integration (WDI) tools are built to deliver scalable, reliable web data extraction out of the box without requiring software development for either configuration or maintenance.

Whereas scraping is a disjointed process of collecting web data that requires software developers to build and reactively maintain custom scraping routines, typically built with a hodgepodge of open source tools.

A modern Web Data Integration (WDI) tool gives you “out of the box” ability to:

- configure web data extraction routines via a user interface with no coding required
- pay for the solution based on results / successful web data extraction
- extract web data across multiple websites and page flows
- merge data from multiple sources
- download multiple file formats (PDF, excel, csv, images, video, audio)
- deliver structured web data via an API or download
- extract your web data behind login screens, two-factor authentication or CAPTCHA



Just as Stripe has done for online payments and Twilio for business communication, modern **Web Data Integration (WDI)** tools are now doing

- automatically adjust pace and frequency to avoid IP blacklisting
- schedule web data extraction (including schedule and return to retrieve)
- ensure data is transmitted and stored securely
- circumvent web scraping defenses
- keep up with quickly evolving web browsers
- validate web data extraction accuracy
- provide audit trails of extracted web data
- leverage experience extracting billions of web data records

Scraping requires hourly, salary and/or contract specialist in product management, software development and QA to build and maintain the same functionality that is available out of the box with Web Data Integration (WDI) tools as listed above.

Plus you to take on the risk of:

- delays or never reaching a production-ready solution
- being out of compliance with data privacy and regulation policies such as GDPR and California Consumer Privacy Act
- instability as you encounter frequent changes and challenges
- cost overruns with building and maintenance
- obsolescence as severity of changes and challenges require a full rebuild
- tying up scarce and valuable software development resources
- having critical IP addresses blacklisted
- insecurely transmitting or storing data
- overwhelming target web data source resources, resulting in blocking
- outsourcing to “gray hat” locales

## To DIY, or Not DIY?

SMB and enterprise adoption of modern Web Data Integration (WDI) tools is on the rise. While there are many legacy, home-grown, outsourced and on-premise solutions still deployed to solve the “web data extraction needs” of competitive businesses, most do not deliver the modern ease of use, reliability and scalability companies need to make web data an integral part of their business.

Consider this: just as Stripe has done for online payments and Twilio for business communication, modern Web Data Integration (WDI) tools are now doing to empower organizations in all markets and of all sizes to capture and act on the endless ocean of web data.

## REFERENCES

1: A. Griem, & O. Marenzi, Web Data Integration — Leveraging the Ultimate Dataset [White paper], retrieved March 2019, Opimas Research