



# The Convergence of Legal + Ethical Scraping

---

How the dos and don'ts of legal and ethical web scraping intersect

# Why care about law and ethics?



**Protect open access to web data** - We firmly believe in open access to web data, but bad actors give the whole industry a bad rep



**Avoid getting sued** - If you follow the law you're less likely to get sued, and if you do get sued you're more likely to win



**Avoid big fines and criminal liability** - GDPR violations may cause hefty fines and CFAA carries criminal penalties, so you want to get these right



**Doing the right thing is just plain cool** - Being a compliant and ethical web scraper protects websites, consumers, and companies that need data. Doing the right thing benefits us all



# Overview

1

## Personal Data

When is it ok to extract personal data?  
What you should and shouldn't do with personal data.

2

## Website Terms

When you need to abide by website terms of service.  
Click Wrap v. Browse Wrap

3

## Copyright

Someone else made something, when is it ok to use that for your own purposes?

4

## CFAA

Updates on the Computer Fraud & Abuse Act in relation to web scraping.

5


## Residential IPs

How to ensure you are using properly sourced IPs.  
How to know if your IPs are ethical?

6

## Cease & Desist

What to do if you get an abuse report or C&D.  
Allow target sites to contact you with any issues.





# Personal Data

## Legality

- Relevant Laws: GDPR + CCPA, with additional US state regulations coming soon
- Ensure lawful basis to scrape – either consent or legitimate interest
- If legitimate interest, do an LIA
- With any lawful basis, notice to data subjects is required
- Right to opt out, deletion, etc is mandatory
- GDPR applies to public data too



## Ethics

- Don't use people's data unless they want you to, either with consent or your use is something they intended or would want
- Respect people's wishes about their data, if they change their mind or want it deleted, honor that
- Don't monetize people's data unless you're selling it for a purpose that would benefit them or they would have intended

# Website Terms

## Legality

- Relevant Laws: contract laws and case law
- Browse Wrap v. Click Wrap: Case law suggests that Click Wrap terms are binding contracts that you must follow
- So if you login to a site, click on agree to terms, then you need to read and follow those terms to the tee

## Ethics

- If you explicitly agree to something, you should follow that
- If information is gated in some manner, ie non-public data, you should respect that it's been gated for a reason and look to the website's limitations around how and when to use that data

# Copyright

## Legality

- Relevant Laws: DMCA in US and Copyright Directive in EU
- If information you scrape is copyrighted, you need to ensure that your use of the data is not infringing on the owner's copyright
- Some information online isn't copyrighted, such as facts, but when it is you either should not scrape or ensure that your use falls under an exception such as fair use
- Fair Use: Fact specific analysis, but think about non-competition with source and transforming the data

## Ethics

Don't copy other people's creative works unless ...

- ✓ You have permission
- ✓ You aren't using it for a competitive purpose
- ✓ You are not republishing it
- ✓ You've transformed it in some way to make it your own work
- ✓ Respect that someone has put time and effort into their work and you should not degrade that

# Computer Fraud & Abuse Act

## Legality

- Relevant Laws: CFAA is a US anti-hacking law that carries civil and criminal penalties
- CFAA prohibits unauthorized access to a computer system
- Recent Case Law: The Supreme Court in Van Buren decided that if you have lawful access to a system but use that system for improper purposes that doesn't violate CFAA, because that's not hacking

## Ethics

- Do not obtain access to a system that you should not rightfully have access to
- Everyone has access to public websites, so it is acceptable to access those systems
- For websites that aren't public, ensure that your access is appropriate and permitted before scraping

# CFAA Case Law

## Van Buren

- Police officer had lawful access to the police database
- He used that access to obtain information that he should not have
- He was sued under CFAA
- The Supreme Court decided that CFAA doesn't apply where you have rightful access to the system to begin with, even if you use that access for improper purposes
- This is a big win for web scrapers, because it tells us that since we have rightful access to public websites we won't be held liable under CFAA
- The DOJ has since released guidance that states it won't pursue criminal action under CFAA for scraping public sites even if terms prohibit it

## LinkedIn v. HiQ

- LinkedIn sued HiQ for scraping its data and sought to have HiQ stop scraping its site during the lawsuit
- HiQ filed an injunction to stop LinkedIn from stopping them from scraping its site during the lawsuit
- The court granted HiQ's injunction, stating that LinkedIn's likelihood of success was low and the damage to HiQ was high
- LinkedIn appealed and the appellate court affirmed
- The case is now back with the lower court for final determination
- Note: The injunction has now been vacated because HiQ is no longer running its business





# Residential IPs

## Legality

- Relevant Laws: Various data protection laws, ie GDPR + CFAA
- Residential IPs are tied to an individual person and are often considered personal data
- Since they are personal data, they need to be obtained in a manner that is consistent with GDPR
- This requires explicit consent from the IP owner and unless consent is expressly and clearly given, it is not lawful



## Ethics

- Do not use a person's IP address without their knowledge and permission
- A person's IP address linked to their computer and data – they have the right to make choices about who they share that with and for what purpose
- Residential IPs are easier to use for abusive purposes, ensure your use is ethical and if selling them conduct KYC checks to ensure proper use

# Cease & Desist

## Legality

- Relevant Laws: case law
- Case law suggests that if you receive a cease and desist you are now on notice of that the target site does not want you to use their data in a specific manner
- Such notice without action can cause legal problems
- If you receive a cease and desist, make sure you have a lawyer respond to it

## Ethics

- If someone doesn't want you to use the data they own, you should respect this where reasonable
- If you still need the data for a legitimate purpose, you should engage with the target site to come to a mutual agreement
- If your use is lawful and they still don't want you to use it, do your best to explain why your use is appropriate – get a lawyers help with this



# Other Considerations



## Legality

- Rate Limits
- Global Data Protection
- Health Data
- Illegal Content
- Adult Content
- Ability to get abuse reports to you

## Ethics

- Do not interfere with the website's operations
- Do not use people's data in a manner that is not consistent with how they would want it to be used
- Do not collect sensitive content unless there is a valid ethical reason

# Resources



---

## Zyte

Zyte has various blog posts, webinars, and panel discussions regarding compliant web scraping



---

## FISD

Web scraping guidelines created by industry experts in the legal and alternative financial data space



---

## i2Coalition

Web scraping companies putting together web scraping guidelines that will set the industry best practices and advocate for fair regulations

# Questions

---

**Sanaea Daruwalla**

Chief Legal Officer

sanaea@zyte.com





Thank you