

# Ethical and legal considerations of web scraping

Web Scraping to build social research data

# Overview

Two key, interrelated questions:

Can you use web scraping in this instance?

Should you use web scraping in this instance?

# Legal considerations

There is a lack of certainty around the legal basis of web-scraping in the UK.

Web scraping may contravene the Terms of Service/Terms of Use/User Agreement of a website.

Web scraping is a means to an end: gathering useful data.

# Ethical considerations

Web scraping is a small part of a larger project which will almost certainly require ethical approval.

Is it ok to deceive the website by masking your identity? By pretending you are a web browser?

What is the impact of your web scraping activities on the website itself (and ultimately its users)?

- Each request you make to a website consumes computational resources, on your end and theirs;
- Overload a server by making too many requests, causing the website to crash;
- Individuals and organisations may rely on a website for vital and timely information, and causing a website to crash could carry significant real-world implications.

# Guidance

Seek ethical approval from institution for research project more broadly.

Read and respect Terms of Service/Use/Agreement and *robots.txt* file.

Seek written permission from website for web scraping activities.

Clearly identify yourself when requesting web pages (e.g., {'User-Agent': 'Dr Diarmuid McDonnell, [d.mcdonnell.1@bham.ac.uk](mailto:d.mcdonnell.1@bham.ac.uk)'}).

Try not to circumvent technical barriers in place (e.g., by masking/rotating your IP address).

Just because the data are public does not mean they are yours! (Check licenses)

# QUESTIONS

# Task

You are interested in collecting information on an ongoing appeal for donations by Great Ormond Street Hospital Charity

<https://www.gosh.org/donate-our-appeal/>

Answer the following questions:

- Can you use web scraping to collect information relating to this appeal?
- Should you use web scraping (or any other method) to collect information relating to this appeal?