

# Process Description – LunchList

After making a shopping list where you have to type out your own products, you get to the problem where misspelling could lead to duplicate products. To prevent this the idea was to link the app to a supermarket API. The problem we are facing is that there are no public supermarket API for the supermarkets in question. This document explores our other options.

## Data scraping

As an alternative to using an API we thought of data scraping, but before we get started with that we need to know if we are allowed to do so.

### The goal

The goal is to collect product information from all the items sold at supermarkets like Albert Hein, Jumbo and Plus. We want a product name, price and image. Maybe we need more in the future, but this is the minimum to start with.

### Legal?

Is it legal to collect this data?

We did some Literature study to investigate this question.

Our first two sources (Source-1, Source-2) tell us that just because the information is publicly available does not mean that it is free to use.

Source-1 mentions “databankenrechten”, this law states the following:

“Kort gezegd moet er een:

Substantiële investering zijn geweest in:

- het verzamelen,
- controleren of
- presenteren van de inhoud van de databank.”

This means if there is a substantial amount of effort poured into collecting, controlling or presenting the contents of the databank it is not available for free use.

We can assume that this is the case with the data of the supermarkets.

Source-2 states that there is an exception. If the scraping is privately used for a project that only gets shared with some friends for example, it is fine.

We also asked ChatGPT about this problem. It noted that these supermarkets probably prohibit scraping in their terms of service. But we could get away with data scraping if

we only collect “small” amounts without stressing the servers too much. This way we avoid getting IP banned or receiving legal claims.

I like to call this “It’s not illegal if you don’t get spotted”.

## Source

Source-1	WiseMen Advocaten
Date	18-3-2025
Link	<a href="https://www.wisemen.nl/nl/artikelen/mag-ik-data-scrapen-van-het-internet-/">https://www.wisemen.nl/nl/artikelen/mag-ik-data-scrapen-van-het-internet-/</a>

Source-2	AP / Autoriteit Persoonsgegevens
Date	18-3-2025
Link	<a href="https://autoriteitpersoonsgegevens.nl/actueel/ap-scraping-bijna-altijd-illegaal">https://autoriteitpersoonsgegevens.nl/actueel/ap-scraping-bijna-altijd-illegaal</a>

Source-3	ChatGPT
Date	18-3-2025
Link	<a href="https://chatgpt.com/share/67d95128-20f4-8001-8250-593f890e2618">https://chatgpt.com/share/67d95128-20f4-8001-8250-593f890e2618</a>