Bachelor thesis
Computing Science

Radboud University

The Extent of Data Gathering by Third Parties on Dutch Health Websites

Author:
Anda Zāle
s1020873

First supervisor/assessor:
dr. Gunes Acar
g.acar@cs.ru.nl

Second assessor:
prof. dr. Erik Poll
erikpoll@cs.ru.nl

January 8, 2022

Abstract

Online privacy has been a widely discussed topic, particularly since the implementation of the GDPR in 2018, which regulates personal data usage. In this thesis we study 178 Dutch health websites from ten different categories in order to investigate data collection by third parties. In addition to in-depth manual traffic inspection, we use DuckDuckGo's Tracker Radar Collector web crawler and Python to automate the process of obtaining the list of third parties present on the websites. Using the combination of manual and automated method, we discover that the median number of third parties vary from 2.0 to 9.0 depending on the website category, with the median being 4.0 across all websites. Furthermore, our results indicate that Google is present on 89.3% of the websites, with google-analytics.com being the most prevalent domain, appearing on 64.6% of the websites. We conclude that while there are potential privacy risks on Dutch health websites, such as mouse tracking and birth date sharing, there are also promising alternatives that are already in use, such as first-party analytics.

Contents

Chapter 1

Introduction

One method of monitoring users' behavior and preferences over the Internet is to utilize online tracking and data collection. Together it could lead to the disclosure of sensitive personal data, among which are various medical conditions. User activity could be followed across websites to identify which topics were selected, as well as the data given in input fields or submission forms, making health-related websites a potential rich source of information. In 2015, a research undertaken by the Electronic Frontier Foundation (EFF) revealed that data collection was present on US healthcare websites[4][20]. Collection of health data is a problem because information of this type could be misused by third parties. As a result, we have focused our attention on Dutch health websites to check for ourselves if and what kind of data third parties collect.

The lack of user understanding about online tracking and data collection is the main source of concern. Data can be coupled with other information gathered across the Web to construct a behavioral profile of a user, which can then be used for targeted advertising, for example. Health-related information, no matter how insignificant, should be treated as personal data and should not be acquired by third parties.

In this thesis, we would like to gain insight on the Dutch health websites to determine whether or not sensitive data collection is a current issue and how severe it is. We look into several categories of websites and use a web crawler to see the type of third parties present on each of them. Manual traffic analysis is used in conjunction with web crawling to determine exactly what data is being collected.

The proposed solution in this thesis demonstrates what is going on behind the scenes of Dutch health websites. What kind of requests are being sent, to whom are they being sent to, and what kind of information is being collected in them? With such information, we are able to assess whether data collection is a concern on Dutch health websites, and whether or not preventative measures should be considered.

We have the following contributions in our thesis:

- Demonstrate an automated method for obtaining a list of third parties present on a list of websites, together with the types of data they collect through manual traffic analysis.
- Provide insight into the most prevalent third-party domains and companies who own them, as well as what kind of data is being gathered.
- Determine whether the scope of data collection provides cause for worry.

We will first describe the theoretical background on the chosen crawler and manual data inspection in Chapter 2. In Chapter 3 we will discuss the URL list used and the automated web crawling method in the context of the programming language Python. Following that, we will analyze the obtained results in Chapter 4. Finally, we will talk about related work in Chapter 5 and draw our conclusions in Chapter 6.

Chapter 2

Preliminaries

2.1 Tracker Radar Collector

In order to obtain data on the requests being made on given websites, we had to choose a web crawler. DuckDuckGo's Tracker Radar Collector (TRC) is a Puppeteer-based web crawler that is modular and multi-threaded (source code is available at `https://github.com/duckduckgo/tracker-radar-collector`). Puppeteer is a Node library to headlessly control Chrome or Chromium over the DevTools protocol[2]. TRC provides third-party request data for Tracker Radar (a DuckDuckGo product with compendiums on third-party domains and their associated domains), but we are utilizing it to gather data for ourselves in this paper.
After each successful crawl, TRC creates a *.json* file named after the crawled website. This *.json* file contains information on requests, APIs, cookies, and full trace events obtained during the web crawl. In addition, a *metadata.json* file containing crawl and system configuration is created. TRC also includes a post-processing script for the generated *.json* file which parses it and returns a *result.json* file. The *result.json* file contains a list of request URLs and domains, which is what we are interested in. Along with that, it also provides information on most commonly contacted IP addresses, JavaScript APIs, and cookies, but these are outside of the scope of this paper.

2.2 Manual Traffic Inspection

To be able to inspect what kind of information is being collected by third parties, we need to interact with the websites and look at the traffic ourselves. A web crawler can visit many websites efficiently, but we need manual analysis to conduct a more in-depth search. As a result, we are using the built-in Developer Tools in the Firefox browser.
The Developer Tools include a network monitoring function that displays all of the requests made upon page load, how long each one takes, and the

details of each request[6]. Any action on a website triggers the Network Monitor to start monitoring the activity. This means that we will be navigating through pages, thus generating traffic for the Network Monitor.

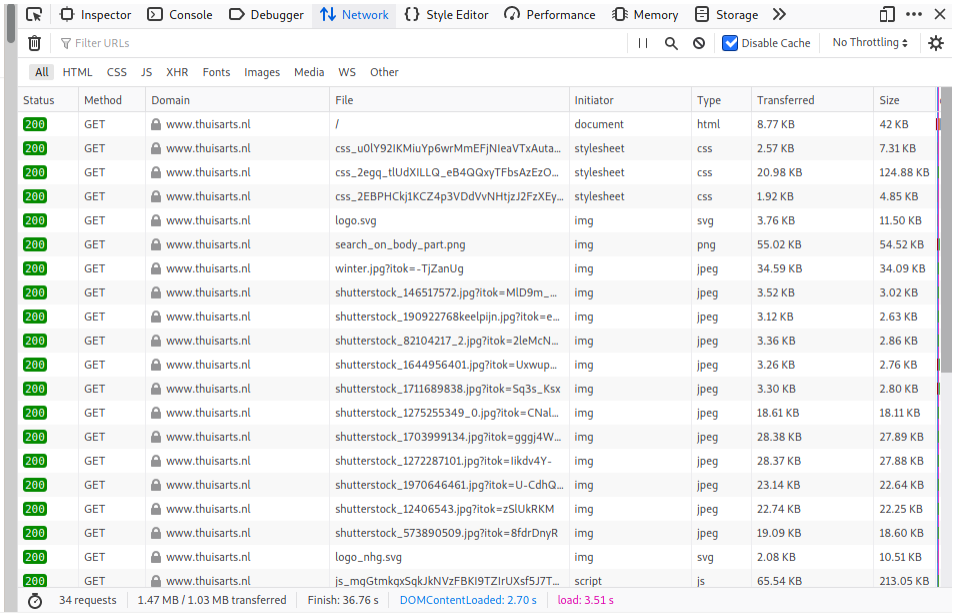

Figure 2.1: *Developer Tools Network Tab.*

We will be looking at the network details window, which contains data like Headers, Requests, Responses, among others. The Headers tab, in particular, contains information about the request, as well as the HTTP response and request headers. The request URL present in this tab contains information that is directly provided to the third party and is unaffected by the referrer policy[7].

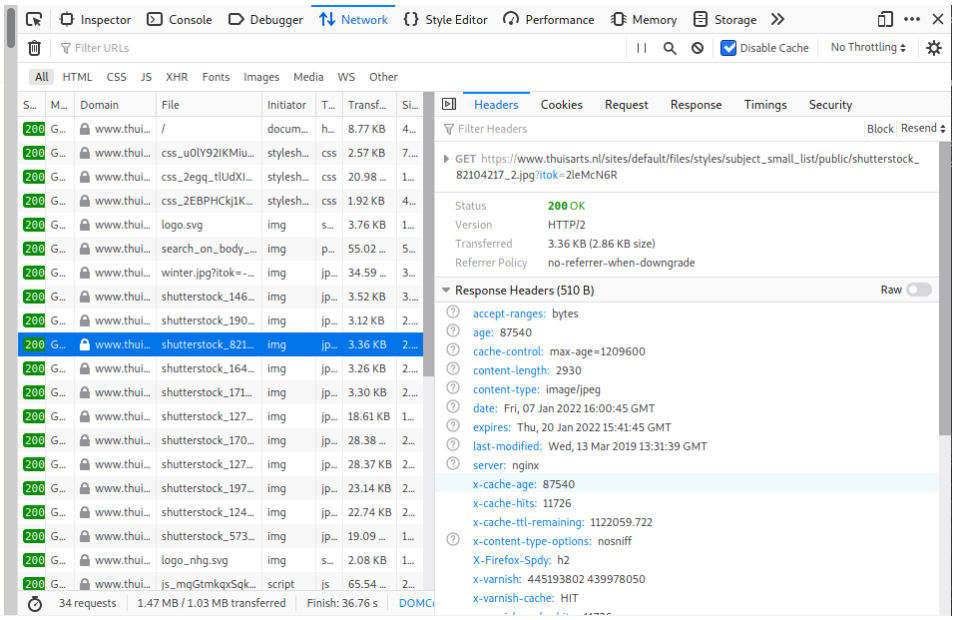

Figure 2.2: *Example of the Details Pane in the Network Monitor.*

Chapter 3

Research

To be able to identify the third parties present on Dutch health websites and assess the data collection performed by them, we will need a list of websites on which to base our thesis on, which will be discussed in Section 3.1. In addition, we will go over the programmed automated method of running the web crawler in Python in Section 3.2, as well as the configuration and policies for manual traffic analysis in Section 3.3. In Section 3.4, we will go over the ethical aspects of the research.

3.1 Data set

While constructing the list of websites to produce data on, we decided to focus on 9 different categories, plus an additional one for when a website fit none of the others. The categories were created with the intention of representing both the governmental and public sectors. We chose all of the available websites whenever possible. When that was not viable, the most popular websites were chosen based on `https://www.zorgkaartnederland.nl/`, which is a website that compiles all Dutch health websites and ranks them on a scale of 1 to 10 based on user reviews. We chose websites with a score of 9.0 or higher. We compiled a total of 178 websites.
The chosen categories for the data set are:

- Governmental
- Health Insurance Companies
- Top-Clinical Hospitals
- University Hospitals
- Online Pharmacies
- General Practitioners
- Dentists
- Physiotherapists
- Health Institutes
- Other

For General Practitioners, Dentists and Physiotherapists, the top rated web-

sites were chosen (rating ≥ 9.0), for the rest all the available ones. The full list can be viewed in the Appendix chapter A.1.

3.2 Automated Data Collection

We need to automate the web crawling process in order to acquire the third parties present on each of the previously selected 178 websites. The automation ensures that we can simply provide a *.txt* file with a URL in each line, and the program will run the web crawler on each website separately, perform the post-processing script, and parse the results into a database. Afterwards, we are able to see all third-party data and compile it to obtain statistics using this new database.
We need to know which domains are third parties and which companies they belong to before we can run TRC on our list of websites. As a result, we employ DuckDuckGo's Tracker Radar. It is a collection of the most common third-party domains on the internet[8]. In particular, we are using their entities folder. The entities folder contains a *.json* file for each company, which lists their associated parameters, including their domains.

Owner	The name of another entity that owns this entity if any
Properties	A list of domains owned by this entity
Name	Name of the entity
Prevalence	The percent of sites this entity is found on

Figure 3.1: *Entity Properties from DuckDuckGo's Tracker Radar[8].*

The Tracker Radar entities folder is parsed into a Pandas dataframe. It is achieved by creating a new Pandas entry for each *.json* file in the entities folder. Only the Properties section of the corresponding *.json* file is extracted for each entry. Each unique Owner-Domain tuple is placed in a separate entry to avoid list column values:

```
import json
import glob
import pandas as pd

jsons_data = pd.DataFrame(columns=['properties'])
for index, js in enumerate(glob.glob("entities/*.json")):
    with open(js, encoding="utf-8") as json_file:
        json_text = json.load(json_file)
        # Extracting the domains for the entity
        properties = json_text['properties']
        # Removing unnecessary information for readability
        jsons_data.loc[js.replace('.json', '').replace('entities\\', '')] = [properties]
# Turn the list column values into separate row entries
jsons_data = jsons_data.explode('properties')
jsons_data.to_csv('jsons_data.csv', index=True)
```

Figure 3.2: *Creating a Dataframe with Companies and Associated Domains.*

We have now created a Pandas dataframe containing companies and their associated domains, which we will refer to as the *Companies and Domains* dataframe. Following on from that, we can use the *Companies and Domains* dataframe in conjunction with the TRC to generate new Pandas dataframes for each of the individual website categories.
To automate the web crawling process, we must write a script that performs the crawl on each website separately. If that is not the case, the web crawl data is combined, making it impossible to determine which results belong to which website. As a result, we wrote a bash script that launches the TRC and runs it on a specified website. The bash script is then used in a **for** loop which iterates through each website in a given *.txt* file.

```
#!/bin/bash
cd
cd PycharmProjects/TrackerFinder/tracker-radar-collector
npm run crawl -- -u $1 -o ../ -v -f
node ./post-processing/summary.js -we ../ -o ../Results/result.json
```

Figure 3.3: *Instructions to Run the Web Crawler on a Given Website.*

After performing the post-processing on a given website, we obtain the eTLD+1s (effective Top-Level Domains + part of the domain just before it) from the domains contacted during the web crawl. In addition, we retrieve the matching domain entries from the *Companies and Domains* database.

```
# Extracting top-level domains from the topDomains list obtained from the crawl
top_domains = [get_fld(item[0], fix_protocol=True) for item in data['requests']['topDomains']]

# Extract the rows from dataframe which were the top domains from the crawl
results = jsons_data.loc[jsons_data['properties'].isin(top_domains)].values.tolist()
```

Figure 3.4: *Extracting the Matching eTLDs+1 from the Companies and Associated Domains Database.*

Finally, we add the web crawl results to a Pandas series that contain the website URL as a column value and the third-party company and domain pairs as row values. The series is then merged with the dataframe containing the results for other websites within the same category.

```
crawl_result = pd.DataFrame(columns=[top_domains[0]])
crawl_result[top_domains[0]] = pd.Series(results)

domains_result = pd.concat([domains_result, crawl_result], axis=1)
```

Figure 3.5: *Merging the New Results Together With the Previous Ones.*

The process of running the bash script and extracting matching values is repeated for each website category, yielding ten distinct dataframes (one for each category).

	efarma.nl	deonlinedrogist.nl	drogisterij.net
0	['Google LLC', 'doubleclick.net']	['Google LLC', 'google-analytics.com']	['Google LLC', 'doubleclick.net']
1	['Google LLC', 'google-analytics.com']	['Google LLC', 'googletagmanager.com']	['Google LLC', 'google-analytics.com']
2	['Google LLC', 'google.com']	['Google LLC', 'gstatic.com']	['Google LLC', 'google.com']
3	['Google LLC', 'google.nl']	['New Relic', 'newrelic.com']	['Google LLC', 'google.nl']
4	['Google LLC', 'googletagmanager.com']	['New Relic', 'nr-data.net']	['Google LLC', 'googleadservices.com']
5	['Google LLC', 'gstatic.com']	['Prospect One', 'jsdelivr.net']	['Google LLC', 'googletagmanager.com']
6	['ImageKit Private Limited', 'imagekit.io']		['Microsoft Corporation', 'bing.com']
7			['Prismic.io Inc.', 'prismic.io']
8			['Zendesk, Inc.', 'zdassets.com']
9			['Zendesk, Inc.', 'zendesk.com']
10			['Zendesk, Inc.', 'zopim.com']
11			
12			
13			

Figure 3.6: *Excerpt from the Online Pharmacies Resulting Dataframe.*

3.3 Manual Data Collection

Now that we have identified the third parties and their associated domains on each website, we need to determine what kind of data they are collecting. To achieve the goal, we interact with the websites directly before inspecting the traffic with the Firefox Developer Tools Network tab.
In order to acquire objective findings, we had to design a protocol for each website category. What the protocol implies is that you should have a broad plan for each website, as in, which parts of the website to navigate to, or whether to check for forms or login fields. The protocols can be divided into several categories:

- For **Health Insurance Companies** the application for health insurance was chosen in each of them. The package and available add-ons were then selected. Following that, we filled out the form that was provided. We used synthetic input rather than personal information to fill out these forms, such as namelastname@gmail.com for *email* or Name and LastName for the *Name* and *Last Name* input fields.

- When it came to **Governmental** and **Health Institutes**, we navigated to the website's 'topics' section. We then chose the more sensitive available categories, ones that a user might not want to share with anyone else, such as cancer, alcohol, or tuberculosis.

- For **Dentists**, **General Practitioners** and **Physiotherapists** we first navigated to the login screen and tried to fill in username and password input fields. Following that, we proceeded to the intake form and filled it out in a similar manner as for the Health Insurance Companies. Finally, if a form for receiving user questions was available, we interacted with it.

- For **Top-Clinical Hospitals** and **University Hospitals** we filled out the intake form in similar manner to the other forms, if one was available. We then attempted to log into their portal. Finally, as with Governmental and Health Institutes, we went to the website's 'topics' section and selected the more sensitive available categories, such as cancer or gynecologist.

- For **Online Pharmacies** the process was simple. We navigated to a urine self-test product and added it to our cart. Following that, we did the same procedure with nicotine patches.

- Finally, because each website is different, the approach for the **Miscellaneous** category was not exactly the same for all the websites. We attempted to interact with any forms provided on the website. Similarly to the other cases, we navigated to the areas of the website that discussed more sensitive themes.

Several measures had to be taken to ensure that the data we collect is representative. To begin, we simulated a typical Internet user by using Firefox's default privacy settings. We also used a Virtual Machine to avoid having to adjust our own browser privacy settings and have them possibly affect our results. Second, before viewing any website, we cleared the web browser's cache and cookies to ensure that we only saw the seed's third parties. If that would not the case, trackers that connect us as users to other websites we have visited may be visible.
Due to the thesis's scope, we limited the inspection to the leading ten websites in each category, and we will examine the results accordingly. Furthermore, if a website made hundreds of requests for each interaction, the protocol was terminated early.

3.4 Ethics

Considering we interacted with multiple websites in this thesis, it is imperative to discuss a few ethical implications. We did not attempt to break the websites throughout the manual traffic inspection process in order to achieve our findings. Moreover, we did not put a lot of load on the website with our automated web crawls. Our goal was to interact with the website as though we were a regular visitor. Furthermore, we made no attempt to collect any kind of personal information about another person. We also did not utilize our own personal information in the forms or input fields, instead opting for synthetic input like 123456789 for phone number.

Chapter 4

Results

In the current chapter we will present our results. In Section 4.1, we will first turn our attention regarding the type of third-party companies and their associated domains present on the selected websites. In Section 4.2 we will discuss the kind of data collected by these third parties. Finally, we will discuss our presented results in Section 4.3.

4.1 Web Crawl Results

We chose to present our obtained results in terms of various statistical data in order to visualize them. We display the median and the average number of third parties present on the websites. In addition, we have graphs depicting the ten most prevalent third-party domains and companies, as well as a histogram of the number of unique third parties.
All websites(n=178). The median is 4.0 and the average number of third parties present is 5.1. Some of the websites we ran the crawler on did not utilize any third parties at all.

Categories	Median of third parties	Average Number of third parties
All Websites	4.0	5.1
Governmental	2.0	2.0
Health Insurance	7.0	7.8
Top-Clinical Hospitals	3.0	3.0
University Hospitals	3.5	3.6
Online Pharmacies	7.0	7.6
General Practitioners	4.0	4.6
Dentists	4.0	4.3
Physiotherapists	5.0	5.4
Health Institutes	9.0	9.0
Miscellaneous	5.0	5.2

Table 4.1: *Website Categories and their Corresponding Median and Average Number of third parties.*

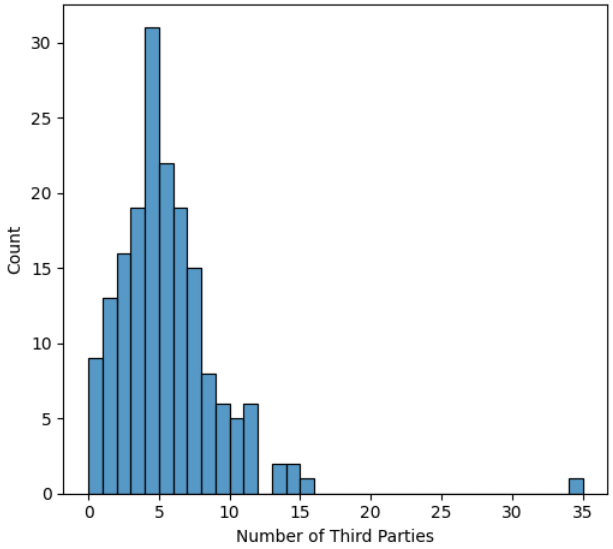

Figure 4.1: *The Histogram of the Amount of Third Parties on All Websites.*

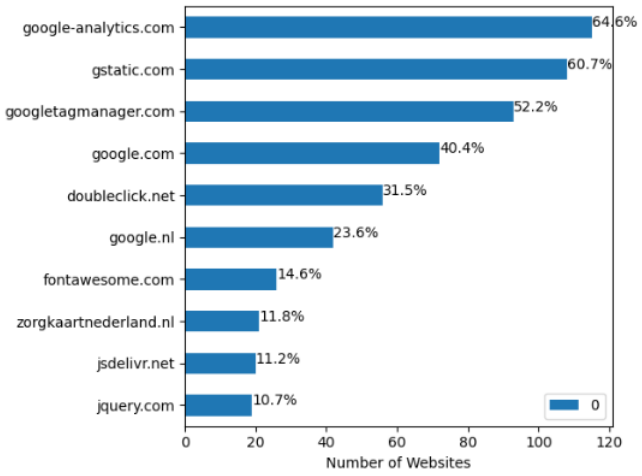

Figure 4.2: *Top 10 Domains on All Websites.*

Figure 4.2 demonstrates that *Google* is represented by the top six domains across the entire website sample, with *google-analytics.com* being the most popular. Nonetheless, *google-analytics.com* is present on 64.6% of the overall sample, which is the typical level when compared to a one million site research, where it was present on approximately 65% of the inspected websites[9].
In seventh place we have *fontawesome.com*, which is a widely-used icon library and toolkit offering full icon customization for websites[22]. *Jsdelivr.net* is a Content Delivery Network (CDN) which loads Internet assets[19]. *Jquery.com* is a JavaScript library for HTML document traversal and manipulation[12].

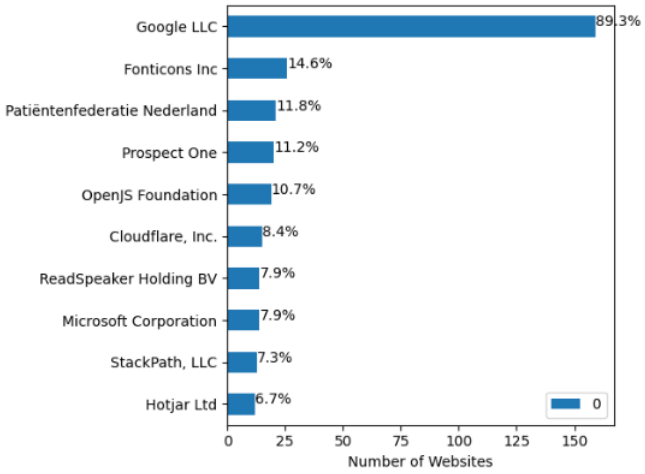

Figure 4.3: *Top 10 Companies on All Websites.*

The top company across all 178 websites is *Google LLC*. Followed by *Fonticons Inc*, in ownership of *fontsawesome.com*. *Patiëntefederatie Nederland* is Netherlands Patients Federation that represents patient organizations to help patients[1]. *Prospect One* is a global digital media lead generation network that assists with traffic generation and advertising. *Prospect One* claims to do that by combining traffic from various sources. They also assist with the monetization of databases and user traffic[29]. *Readspeaker Holding BV* owns domains that offer accessibility services to first parties, such as reading out loud text written on a website[30]. *StackPath* offers cloud computing services which are put in densely populated markets[13]. *Hotjar*, which provides website heatmaps, session recordings, and other behavior analytics tools, is ranked tenth[26]. Such methods potentially expose users to privacy threats, such as the collecting of passwords or login information in the sessions recordings via session replays.

Health Insurance Companies (n = 33). We have that the median is 7.0 and the average number of third parties present is 7.8. Both of these statistics are much higher than the global average. It is possible that the high number of third parties is related to the fact that health insurance firms are always looking for new customers to purchase their products.

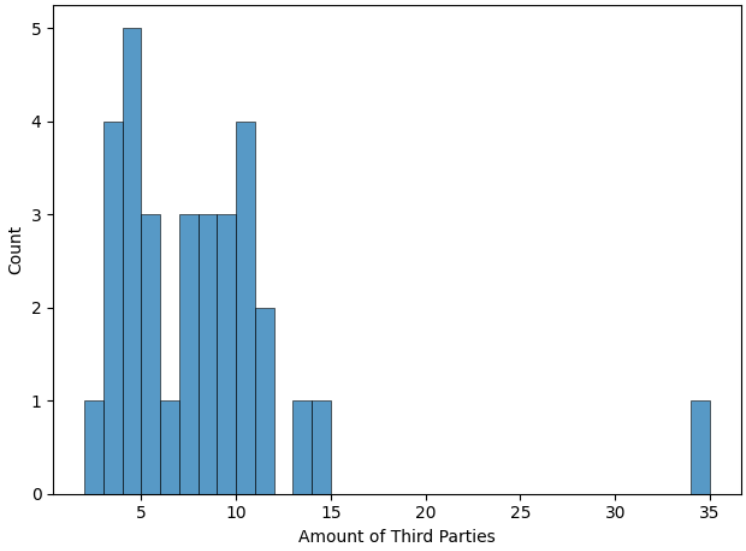

Figure 4.4: *The Histogram of the Amount of Third Parties on Health Insurance Websites.*

As illustrated in Figure 4.4, one of the Health Insurance Company websites is an outlier with 35 unique third parties. This website is `zorgkiezer.nl`.

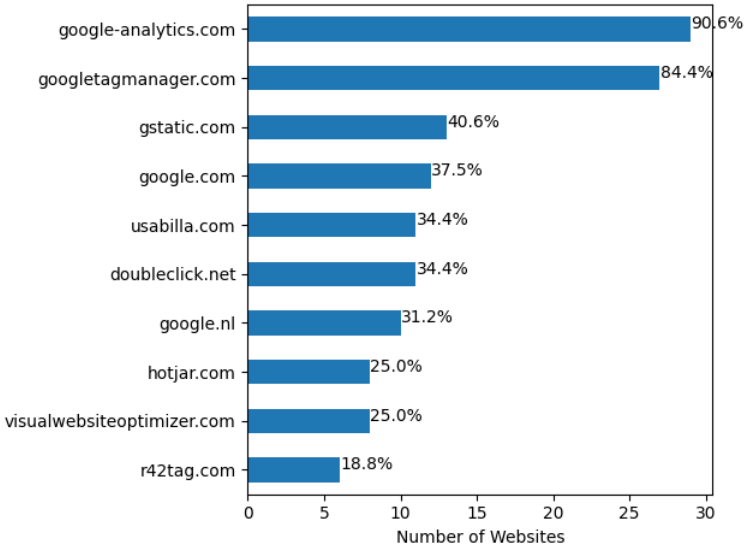

Figure 4.5: *Top 10 Domains on Health Insurance Websites.*

Usabilla.com is an Amsterdam-based service providing real-time feedback to improve customer experience that is used by major corporations such as ABN Amro and KLM[34]. Real-time feedback provides the potential to

capture IP addresses and mouse movements, which could be problematic regarding privacy. *Visualweboptimizer.com* is used for A/B testing (split testing), a method of comparing two versions of a single variable to determine which is more effective[35]. A/B testing is claimed to enhance user experience. *r42tag.com* is a domain that belongs to Relay42, which is a data management firm for "Intelligent Journey Orchestration"[31].

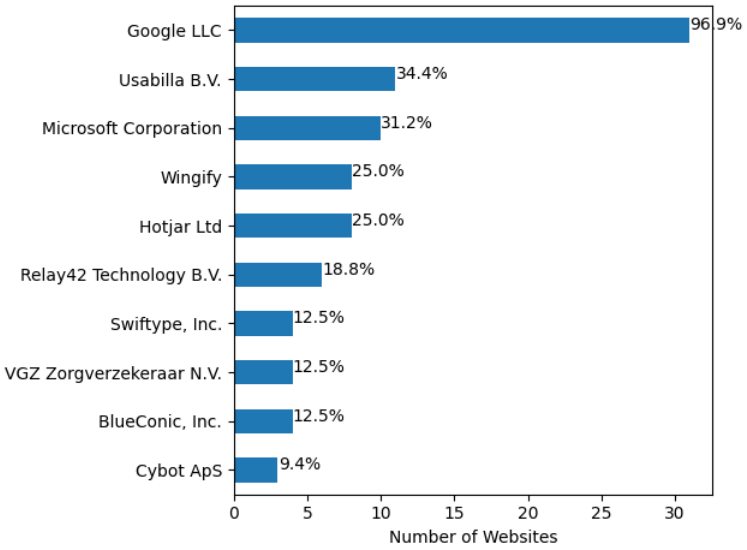

Figure 4.6: *Top 10 Companies on Health Insurance Websites.*

Swiftype, Inc is a search and index company that offers a 'world-class' search that delivers highly relevant results[21]. *VGZ Zorgverzekeraar N. V.* is the Collective Health Insurance[3]. The presence of VGZ on five domains could be attributable to the fact that it is an umbrella organization with numerous domains. *BlueConic, Inc* is a Customer Data Platform that allows to collect data and interact with visitors[23]. *Cybot ApS* delivers automated ePrivacy services, most commonly used for *cookiebot.com*, which is a consent solution that ensures compliance with GDPR, CCPA and other cookie laws[24].

Top-Clinical hospitals (n = 26). The median and average number of third parties present is 3.0.

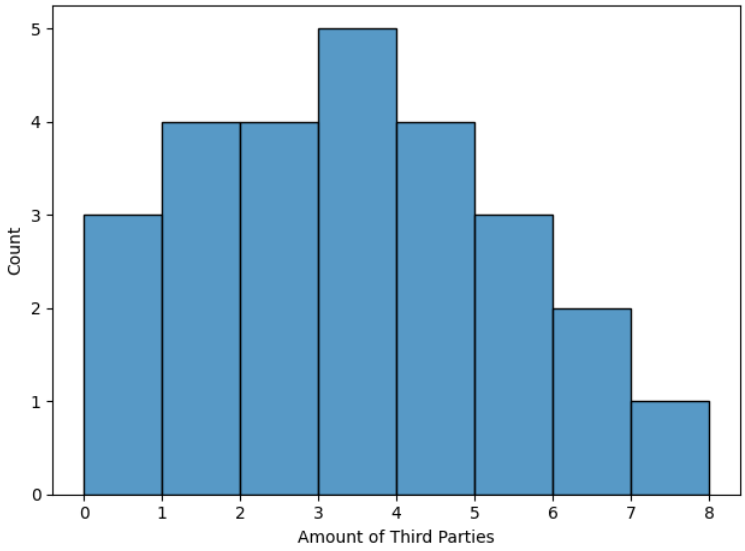

Figure 4.7: *The Histogram of the Amount of Third Parties on Top-Clinical Hospital Websites.*

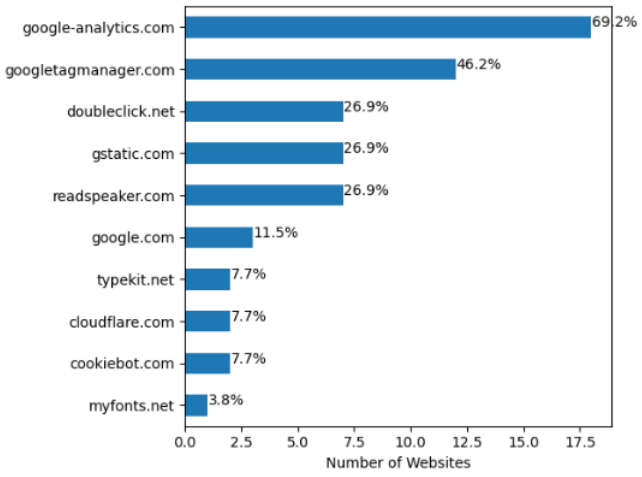

Figure 4.8: *Top 10 Domains on Top-Clinical Hospital Websites.*

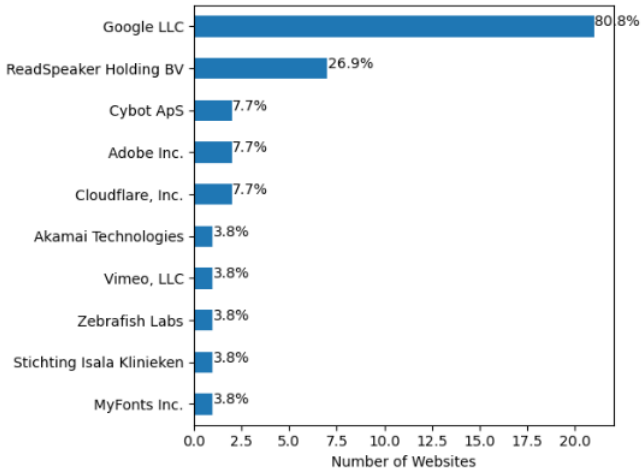

Figure 4.9: *Top 10 Companies on Top-Clinical Hospital Websites.*

Cloudfare.com, owned by *Cloudfare, Inc*, provides content delivery network and DDoS mitigation services[38]. *Zebrafish Labs* offers image processing APIs[36].

University Hospitals (n = 8). The median is 3.5 and the average number of third parties present is 3.6.

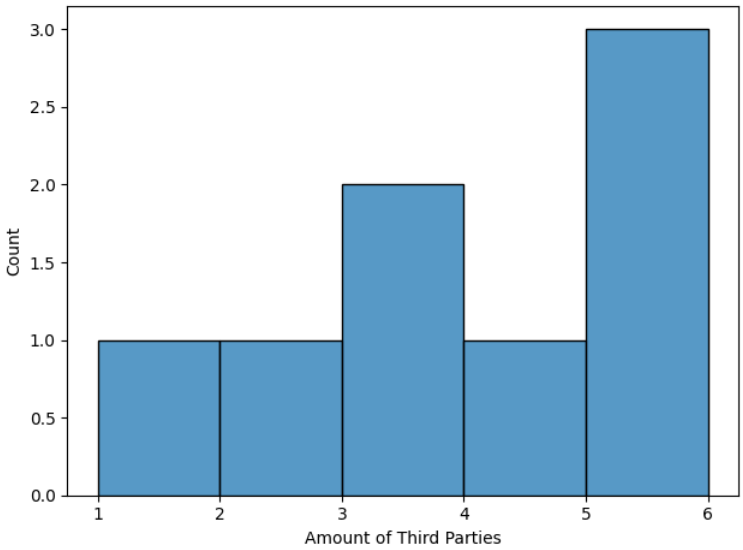

Figure 4.10: *The Histogram of the Amount of Third Parties on University Hospital Websites.*

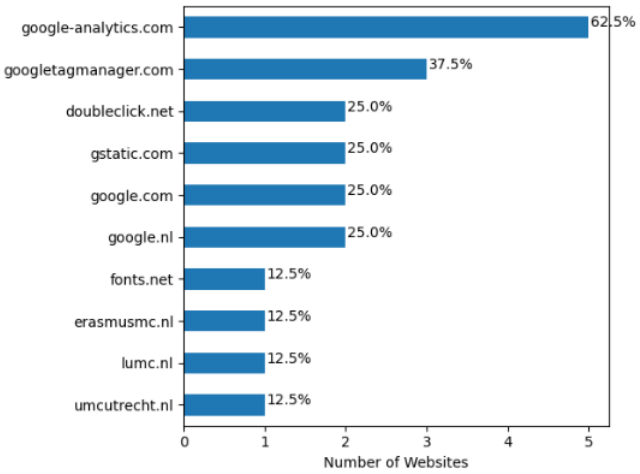

Figure 4.11: *Top 10 Domains on University Hospital Websites.*

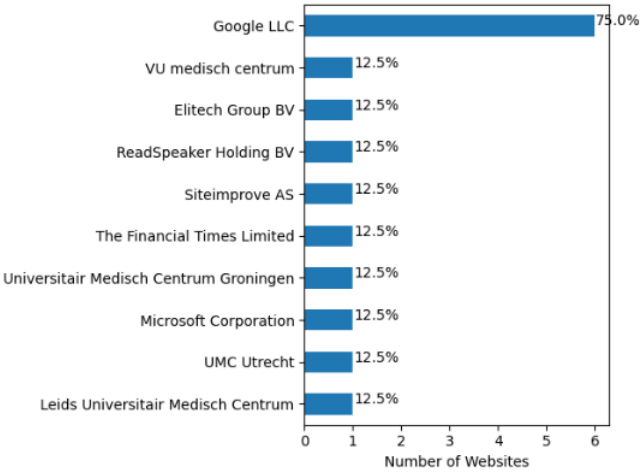

Figure 4.12: *Top 10 Companies on University Hospital Websites.*

The Financial Times Limited is one of the world's leading news organisations, and it also offers a service to provide functionality of websites on older browsers that do not natively support it[25].

Online Pharmacies (n = 8). The median is 7.0 and the average number of third parties present is 7.6, which is twice the amount from the overall statistics. The high number of third parties could as well be attributed to consumerism.

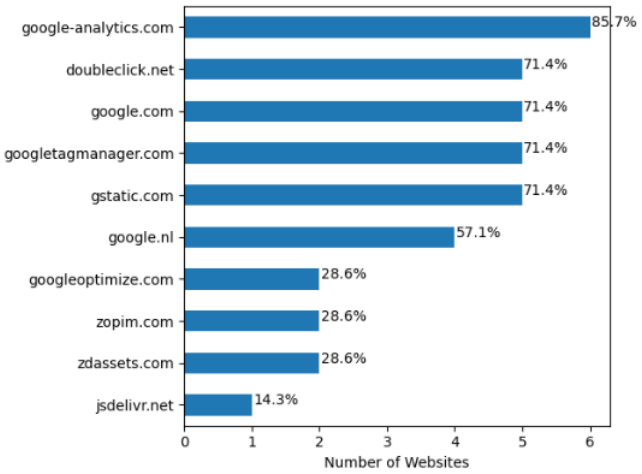

Figure 4.13: *Top 10 Domains on Online Pharmacy Websites.*

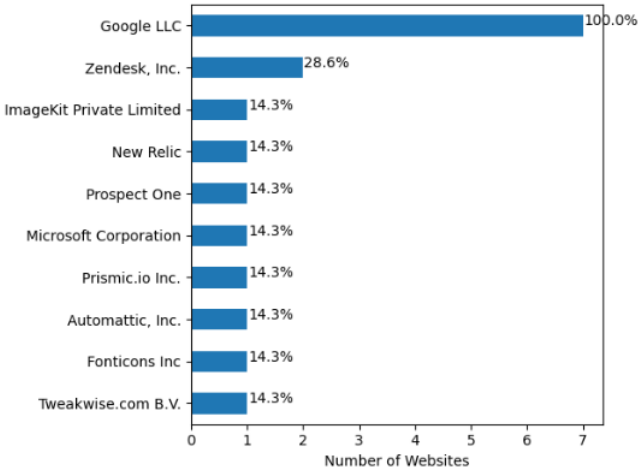

Figure 4.14: *Top 10 Companies on Online Pharmacy Websites.*

Zopim.com and *zdassets.com* are owned by *Zendesk, Inc*, and used for helpdesk chat services[37]. *New Relic* is used for tracking performance of services and software applications[39]. *Tweakwise.com B. V.* is a Netherlands-based service offering personalized shopping experience[33]. *Prismic.io Inc* is a Content Management System (CMS) that allows you to make changes to your online content[28].

General Practitioners (n = 70). The median is 4.0 and the average number of third parties present is 4.6.

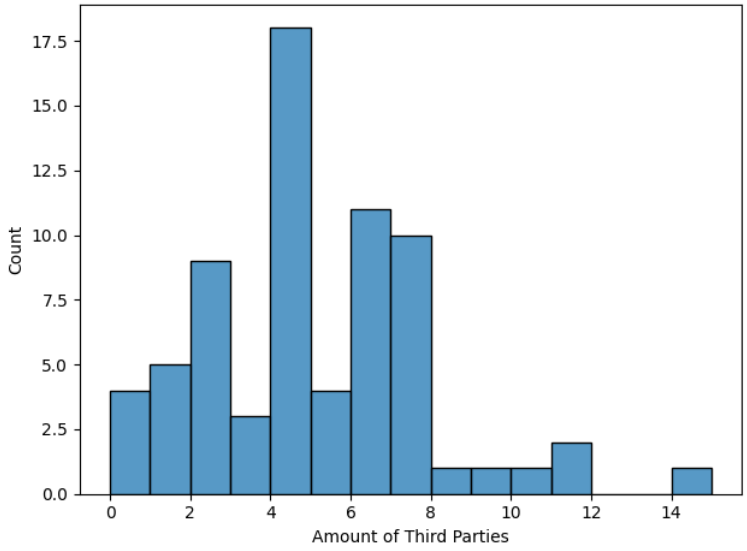

Figure 4.15: *The Histogram of the Amount of Third Parties on General Practitioner Websites.*

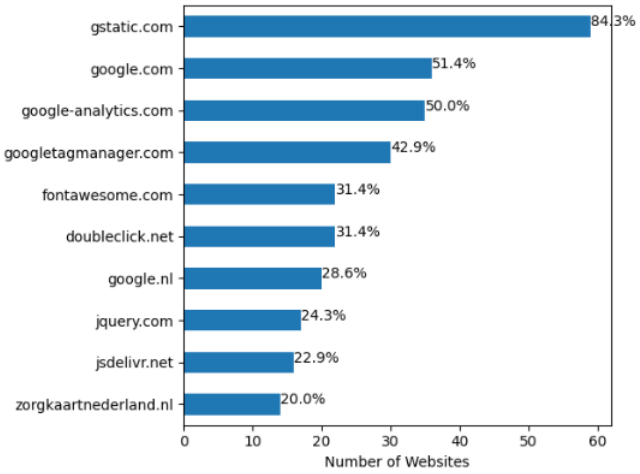

Figure 4.16: *Top 10 Domains on General Practitioner Websites.*

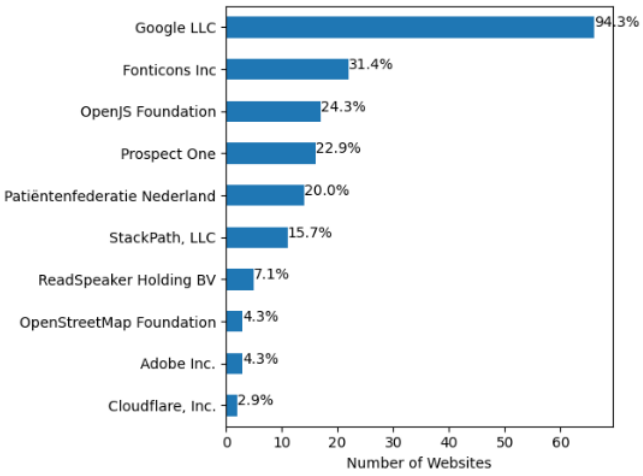

Figure 4.17: *Top 10 Companies on General Practitioner Websites.*

Physiotherapists (n = 9). The median and the average number of third parties is 5.4.

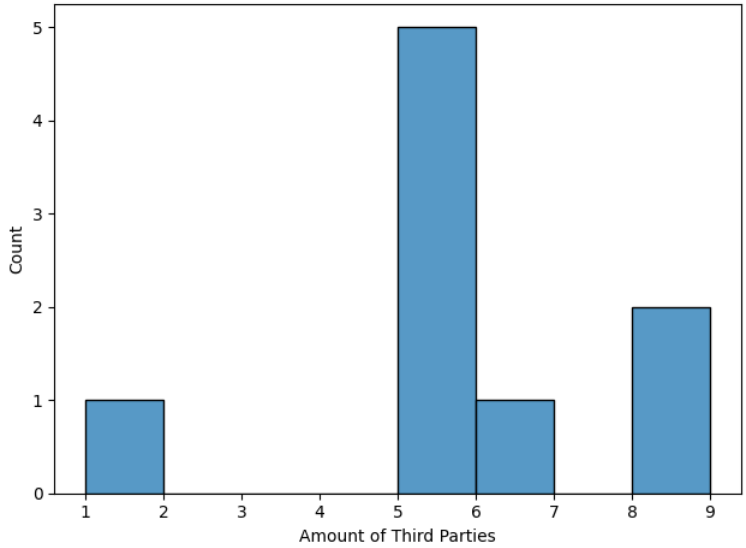

Figure 4.18: *The Histogram of the Amount of Third Parties on Physiotherapist Websites.*

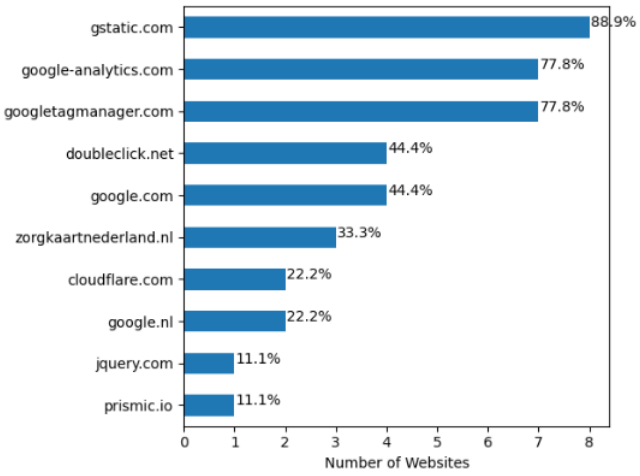

Figure 4.19: *Top 10 Domains on Physiotherapist Websites.*

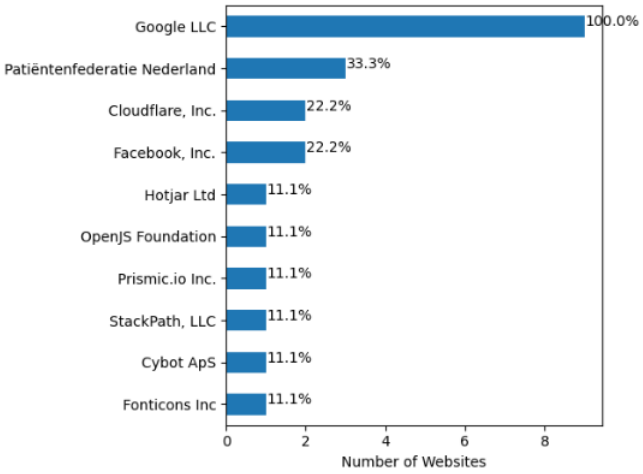

Figure 4.20: *Top 10 Companies on Physiotherapist Websites.*

As depicted in Figure 4.20, *Facebook, Inc* is one of the top ten corporations on physiotherapy websites. This is the company's first appearance from all the previously seen categories.

Dentists (n = 14). The median and the average number of third parties present is 4.3.

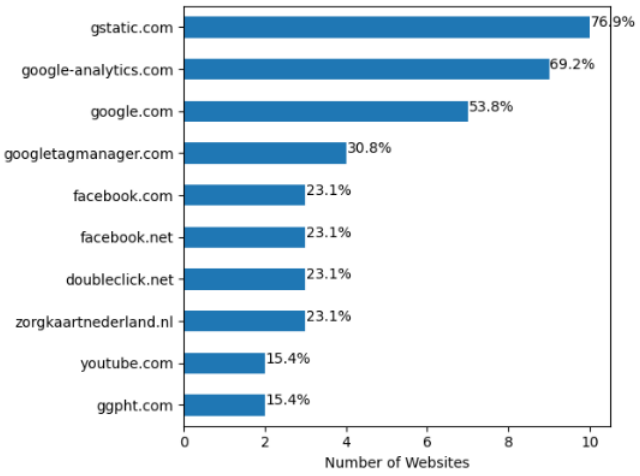

Figure 4.21: *Top 10 Domains on Dentist Websites.*

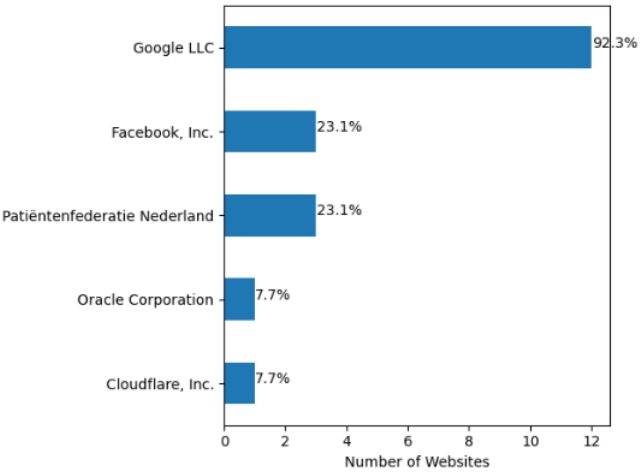

Figure 4.22: *Top 10 Companies on Dentist Websites.*

Figure 4.22 shows that *Facebook, Inc* is the number two company among dentist websites. It is also one of the most popular domains.

Health Institutes (n = 2). The median and the average number of third parties present is 9.0, which is significantly higher than statistics on all 178 websites.

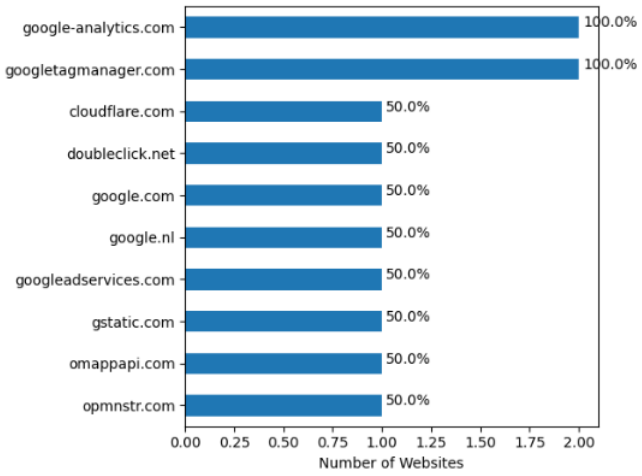

Figure 4.23: *Top 10 Domains on Health Institute Websites.*

omappapi.com and *opmnstr.com* is owned by *Retyp LLC*, or *OptinMonster*, which offers to transform traffic into subscribers, leads and sales. Exit intent recognition, gamified wheels, and other methods are claimed to help them grow email lists[32].

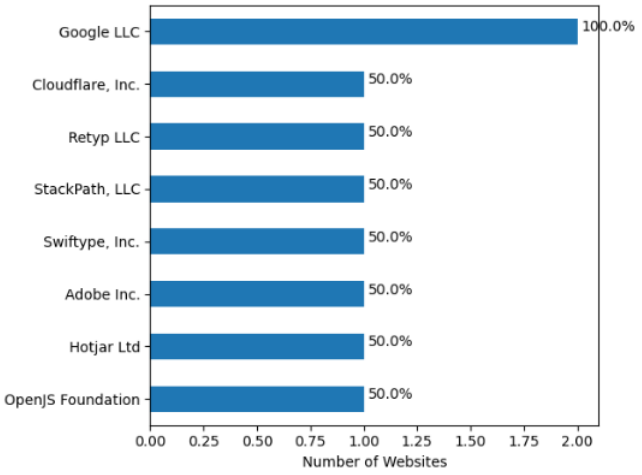

Figure 4.24: *Top 10 Companies on Health Institute Websites.*

Miscellaneous (n = 6). The median is 5.0 and the average number of third parties present is 5.2.

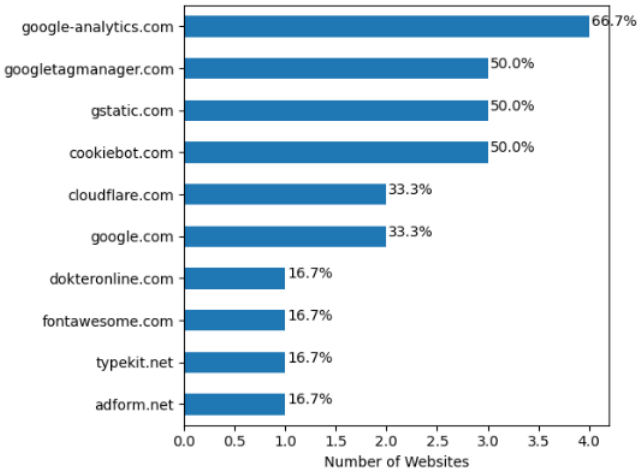

Figure 4.25: *Top 10 Domains on Miscellaneous Websites.*

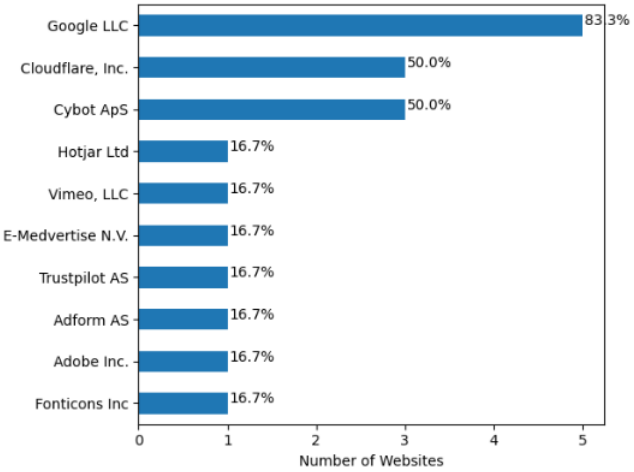

Figure 4.26: *Top 10 Companies on Miscellaneous Websites.*

Governmental (n = 2). The median and average number of third parties present is 2.0, which is significantly lower than the overall average statistics. The decrease could be linked to the lack of competition for government websites, which means they are less concerned with increasing web traffic.

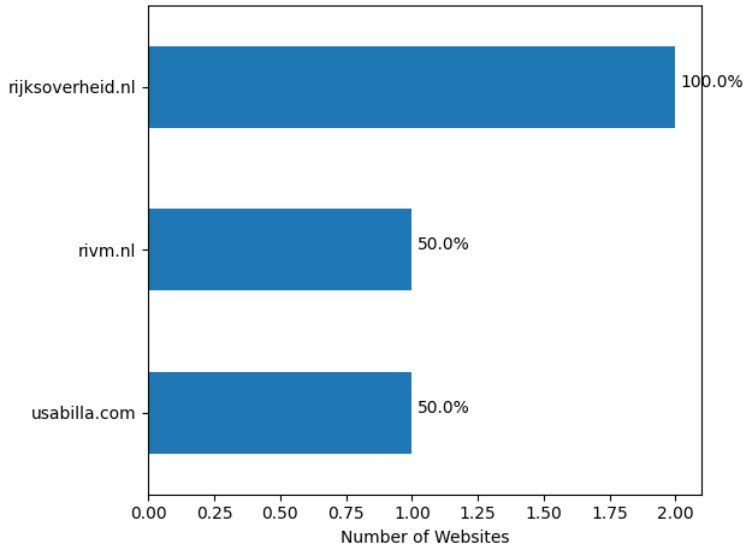

Figure 4.27: *Top 3 Third Party Domains on Governmental Websites.*

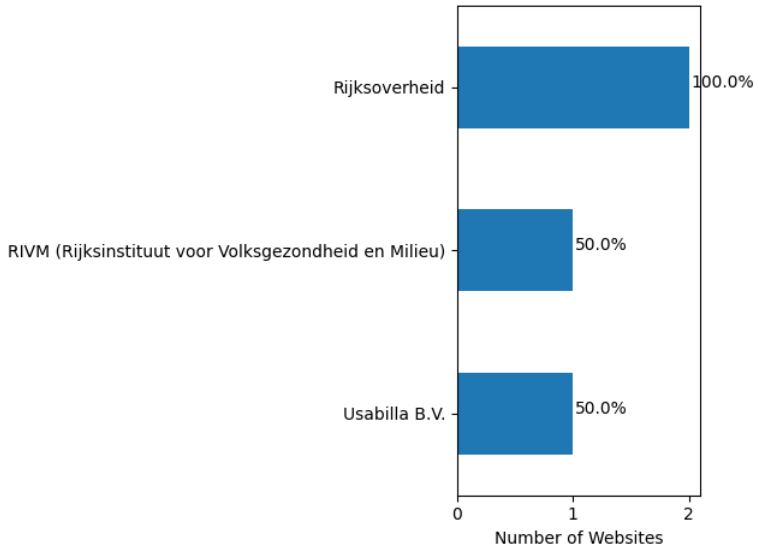

Figure 4.28: *Top 3 Companies on Governmental Websites.*

The Dutch government owns the top two domains and companies in both Figures 4.28 and 4.29. The Central Government and the National Institute for Public Health and the Environment, respectively, are known as *Rijksoverheid* and *RIVM (Rijksinstituut voor Volksgezondheid en Milieu)*.

4.2 Manual Interaction Results

In the current section, we demonstrate the various results we observed during our manual interactions with the websites. We chose examples that illustrate how users' movements are tracked, what type of information is gathered by third parties, or the most representative requests based on the website category.

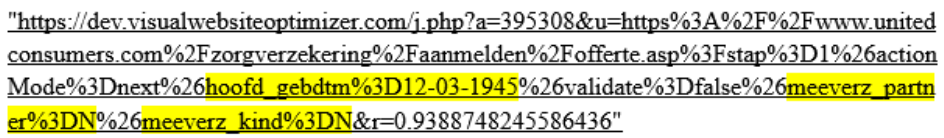

Figure 4.29: *unitedconsumers.nl Request To visualweboptimizer.com.*

Birth date collection. Figure 4.29 depicts one of our discoveries on the website *unitedconsumers.nl* in the Health Insurance Company category: the birth date is provided to the *visualweboptimizer* third party in the request URL. Apart from that, we can observe that they collect information whether the user has a partner or children. This was the only time such a request was observed. As such information is contained in the URL parameter of the request, it is possible that the information was leaked as a result of a first-party error. This type of error can occur when a first party unintentionally includes identifying information in a URL or page title[16].

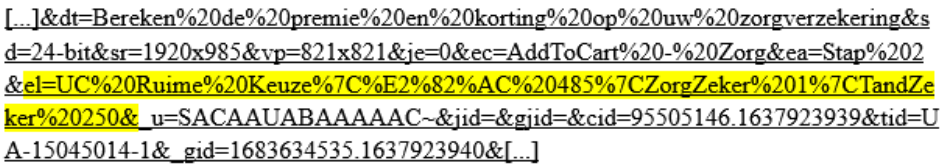

Figure 4.30: *Excerpt of unitedconsumers.nl Request To google-analytics.com.*

Health insurance package details. Figure 4.30 also demonstrates that, in addition to the prior birth date, partner and children data, *google-analytics* obtains specifics about the selected insurance package, such as the type of insurance and extra services. We noticed that collecting health insurance package details was common across the selected Health Insurance websites, with the addition of the selection's price details.

Facebook on application forms. Furthermore, when it came to Health Insurance websites, apart from the previous findings, we saw that

Facebook regularly appeared as a new third party in the form of *Facebook pixel* after navigating to the forms for health insurance application. When a visitor on a website takes an action, the *Facebook pixel* is triggered, allowing the first party to contact the consumer again using Custom Audience[40]. As example, this pixel might be used to create customized adverts based on an existing web profile. Since Facebook was not present on the website's homepage, it did not show up in the web crawl results. An example of a *facebook.com* request is as follows:

"https://www.facebook.com/tr/?id=999824866745760&ev=PageView&dl=https%3A%2F%2Fwww.unitedconsumers.com%2Fzorgverzekering%2F&rl=&if=false&ts=1637924241739&sw=1920&sh=985&v=2.9.48&r=stable&ec=0&o=30&fbp=fb.1.1637924241737.1927788608&it=1637924241345&coo=false&exp=p1&rqm=GET"

Figure 4.31: *unitedconsumers.nl Request To facebook.com.*

[...]\"currentosversion\":[\"Linux Unknown\"],\"mr_geo_zipcode\":[\"6512\"],\"mr_geo_subdivision_2_name\":[\"Gelderland\"],\"firstvisit\":[\"1637923266435\"],\"mr_geo_city_name\":[\"Nijmegen\"],\"sent_to_system\":[\"gtm\",\"ua\"],\"geo_city_name\":[\"Nijmegen\"],\"visiteddomain\":[\"59985613-036e-4a1f-b3cf-5b60b7161c4a\"],\"mr_geo_continent_name\":[\"Europe\"],\"mr_geo_latlong\":[\"51.8421,5.8593\"],[...]

Figure 4.32: *Excerpt of a Response To a blueconic POST Request on salland.nl.*

Geolocation data. A POST request to the third party *blueconic* on *salland.nl* is shown in Figure 4.32. They are acquiring the user's geolocation, in our case Gelderland, Nijmegen, as we can see in the response header. *Blueconic* also attempts to obtain the operating system the user is using, which could be a sign of system fingerprinting.

4{"c":"visitorMessage","p":[{"uid":null,"n":"Jan Ronald","ut":"s","t":"c","m":"Welkom bij Fysio Masters.\nKan ik u helpen?","cver":3,"co":"2021-12-27T12:59:36.635Z","origin":"triggers","md":{"pi":"5d1b07dd22d70e36c2a3b420/var/trigger-images/444c8767c762490e0f4bcf8f0b8a6c386ffe57e0.jpg"},"vsk":"61c9b8899350042b6555eff6"}]}
4{"c":"service","cb":0,"p":["visitor-chat","/v1/visitor/dismiss-preview",{"timestamp":1640609976635}]}

Figure 4.33: *Messages To a websocket on fysiomasters.nl*

Mouse tracking. Figure 4.33 shows a websocket to *vsa85.tawk.to* running on *fysiomasters.nl* (a physiotherapist website) that records where the user

clicks on the website. In this case, it shows how and when the visitor closed a chat pop-up message that appeared just before the event. Session replays may benefit from the retrieval of information such as mouse movement or clicked elements on the website. We also noticed a similar websocket on *interpolis.nl* (a health insurance website), but this time it recorded mouse movement in terms of relative coordinates, as shown in the figure below:

{"uuid":"55d76b48-e12f-58c4-a6c8-b0df6e0b61fe","viewportid":"08546aa3-6d24-4d25-bb18-15243cca5ae8","site_id":552119}
{"mouse_move":[{"time":15714,"timestamp":1640611879667,"x":513,"y":297}],"relative_mouse_move":[{"offset_x":70,"offset_y":6,"selector":"0:#ipoclick-btn-pa-home-mijn_verzekeringen"}],"page_visit_key":"d489639a-f354-494c-8383-df293a68c71c","viewport_id":"08546aa3-6d24-4d25-bb18-15243cca5ae8"}

Figure 4.34: *Message To a websocket on interpolis.nl*

"https://www.google-analytics.com/collect?v=1&_v=j96&a=1484709694&t=event&ni=1&_s=1&dl=https%3A%2F%2Fwww.drogisterij.net%2Fbayer-ketostix-teststrips-voor-urine-5016003288005.html&ul=en-us&de=UTF-8&dt=Drogisterij.net&sd=24-bit&sr=1920x985&vp=760x821[...]&pal=%2Fgezondheid%2Fzelftesten%2Furinetest&pa=detail&pr1nm=Ketostix%20Teststrips%20Voor%20Urine%20*Bestekoop&pr1id=156463&pr1pr=7.62&pr1br=Bayer&pr1ca=Urinetest&z=686194687"

Figure 4.35: *A google-analytics.com request on drogisterij.net*

Product details. Figure 4.35 illustrates a *google-analytics.com* request on the online pharmacy website *drogisterij.net*, which contains product details for the selected product, in this case a urine self test. The type of product, its price, and product id, as well as the product category to which it belongs, are all included in the request. Similar activity has been seen on other online pharmacy websites, such as *deonlinedrogist.nl* and *medicijnen.nl*. The retrieval of the types of products that the user has inspected or chosen has the ability to reveal the user's potential health issues.

"https://www.google-analytics.com/j/collect?v=1&_v=j96&a=2112469296&t=pageview&_s=1&dl=https%3A%2F%2Fwww.lumc.nl%2Forg%2Fgynaecologie%2Fgynaecologische-zorg%2Fpoliklinische-behandelingen%2F&ul=en-us&de=UTF-8&dt=Gynaecologische%20zorg%20in%20het%20LUMC%20%7C%20LUMC&sd=24-bit&sr=1920x985&vp=854x821&je=0&_u=CCCAAAAjAAAAAC~&jid=&gjid=&cid=1533461259.1637930858&tid=UA-34412642-1&_gid=2141286141.1637930858&_slc=1&z=1563203686"

Figure 4.36: *A google-analytics.com request on lumc.nl*

URI, analytics and resolution. The most common *google-analytics.com* request that we encountered on a large number of the chosen websites is shown in Figure 4.36. As can be seen, Google obtains the whole URL of the website the visitor is browsing, in this case the gynecologists subcategory. It also records the current user's page view, as well as the screen and browser resolutions.

"https://statistiek.rijksoverheid.nl/piwik/piwik.php?action_name=Alcohol%20%7C%20Rijksoverheid.nl&idsite=4&rec=1&r=331100&h=4&m=36&s=45&url=https%3A%2F%2Fwww.rijksoverheid.nl%2Fonderwerpen%2Falcohol&urlref=https%3A%2F%2Fwww.rijksoverheid.nl%2F&_id=14b2e474385173b8&_idts=1637746568&_idvc=1&_idn=0&_viewts=1637746568&send_image=1&cookie=1&res=1920x985&[...]%7D&gt_ms=670&t_us=740&t_ue=740&t_ss=0&t_fs=4&t_ds=4&t_cs=4&t_ce=4&t_qs=64&t_as=733&t_ae=734&t_dl=739&t_di=1233&t_ls=1234&t_le=1238&t_dc=1260&t_ee=1264&pv_id=jaVWPb"

Figure 4.37: *rijksoverheid.nl statistics request.*

Self-hosted analytics. Finally, Figure 4.37 shows how the government website *rijksoverheid.nl* collects website statistics from users using **piwik** (now also known as **Matomo**). This is an open-source alternative to Google Analytics, with the difference being that the owners retain complete ownership of the collected data[27]. The statistics the first party collects are which parts of the website a user visits, along with the screen resolution.

4.3 Discussion

We go over our findings and discuss the implications they suggest. In Section 4.1, we showed that the median number of third parties across all 178 websites is 4.0 and the average number of third parties is 5.1. Furthermore, the statistic did not change much among the different categories, with the exception of Health Insurance Companies, Online Pharmacies, and Governmental. The increased median and the average number of third parties present on Health Insurance Companies (7.0) and Online Pharmacies (7.0) could be related to the desire to attract more potential clients, as mentioned

earlier in the chapter. On the other hand, some website categories, such as Governmental (2.0) or Hospitals (3.0 and 3.5), showed a lower number of the median and average number of third parties present. One of the reasons for this could be that their respective fields are not naturally competitive. Second, rather than providing services, their primary goal is to provide information. That could be the case especially for Governmental websites, as they would have no need to monetize traffic.
One limitation of the website selection within each of the different categories is that it makes comparing the findings difficult. Specifically, the number of webpages in each category varies substantially (min: 2, max: 70). Another limitation is that we used the Firefox browser during manual analysis, which includes some built-in privacy measures. Some social media trackers and cross-site tracking cookies are blocked by the built-in privacy, but several tracking content is allowed because it would otherwise break elements of a website[10]. Other browsers, such as Opera, Apple Safari, and Brave, have similar browser privacy settings[18][11][5]. However, using Firefox for manual website interactions has no effect on our web crawling results, as TRC is based on headless Chromium, which does not have any tracker protection.
Nonetheless, based on the findings, we may conclude that on average there are several different third parties present on Dutch health websites, with Google taking the lead (89.3%), but then it quickly drops off (Fonticons with 14.6%). Furthermore, the amount of third parties present on Dutch health websites is substantially smaller than on news websites[9]. In addition, certain third parties, such as *ReadSpeaker Holding* or *Cybot ApS*, provide useful services like accessibility or GDPR/CCPA compliance.
We can see from Section 4.2 that some websites are capable of collecting website analytics using self-hosted solutions like *Matomo*. If the first party's purpose is to acquire data on website analytics, this might potentially eliminate the need for third-party data sharing entirely.
On the other hand, *google-analytics.com* is used on 64.6% of the websites for analytics purposes. As a result, Google may gather more data about users across the web for a variety of purposes, including targeted advertising and online profiling. Furthermore, there are a number of cases where more information is collected than just web analytics, which is concerning. This includes visitor's birth date, geolocation, and mouse movement tracking. Such information could potentially be used by third parties for fingerprinting, session replay, web beacons, tracking pixels, or tracking in general, to mention a few examples.

Chapter 5

Related Work

There have been numerous studies on analyzing web privacy. A notable related study is a paper published in 2015 by Libert on the privacy implications of health information on websites based on searches for common diseases. A total of 80142 web pages were gathered while focusing on 1986 common diseases. The *PhantomJS* browser was used in conjunction with Python in Liberts study to obtain a list of third parties present on these websites.[14]. In our thesis, we were able to streamline the process thanks to the DuckDuckGo Tracker Radar third-party collection, which was released in 2020 and includes more than just trackers, as well as their web crawler TRC (covered in Section 2.2).
In 2018, Libert also released a paper on an automated approach to auditing the disclosure of third-party data collection website privacy policies. With the help of *webxray* and *policyxray*, which are used to determine data collecting third parties and extract privacy policies for auditing, they were able to determine all unique domains per page-load on 938093 websites. One of the findings was that Google tracked over 82% of the scanned websites.[15]. This study did not concentrate on a single category, but rather on the most popular websites as determined by the Alexa company.
Niforatos, Zheutlin, and Sussman recently investigated the prevalence of third-party data tracking by US hospital websites in a 2021 research letter. They collected information on advertisement tracking, third-party cookies, session recording devices, and Facebook and Google tracking software using *Blacklight* (a tool that monitors website surveillance scripts)[17]. Unlike this study, which focused on identifying threats, we investigated all third parties and what type of data was being collected.
In 2016, Englehardt and Narayanan measured and analyzed one million sites. They identified a problem with web privacy measurements and created *OpenWPM* to address it. This platform is capable of automating the simulation of users and the recording of observations[9]. For our research, we considered using *OpenWPM* instead of TRC, but it has one disadvantage:

it only works on Linux operating systems.

Chapter 6

Conclusions

In our thesis we studied the type of third parties present on Dutch health websites, as well as the kind of data they collect. We compiled a list of 178 websites from ten health-related categories. We presented an automated method for obtaining the list of third parties on the chosen websites using Python and DuckDuckGo's TRC and third-party database. In addition, we manually interacted with the websites and examined network traffic through Firefox Developer Tools in an effort to see the type of data requested by third parties.

We were able to show that the median of third parties is 4.0, and the average number of third parties present is 5.1. Furthermore, our analysis show that Google is the most prevalent company on Dutch health websites, with its presence being 89.3% of the websites in our sample. Furthermore, the leading 6 domains, with *google-analytics.com* as first with 64.6%, are all owned by *Google*. We also uncovered that apart from companies dedicated for traffic monetization and advertising, some third parties are used for accessibility, privacy assurance or functionality and design.

During our manual analysis of websites, we uncovered several instances of information leaks. Apart from website analytics and URIs, mouse movement tracking, geolocation, system and browser properties, and birth date was collected. These findings are a cause for concern as it can be misused for attacks on the user or online profiling. Encouragingly, certain websites use first-party analytics solution to further prevent the collection of use data by third parties.

Bibliography

[1] Patiëntenfederatie nederland. https://www.patientenfederatie.nl/. Accessed on 2021-12-10.

[2] Puppeteer. https://pptr.dev/#faq. Accessed on 2021-12-10.

[3] Vgz collective health insurance. https://www.vgz.nl/english. Accessed on 2021-12-10.

[4] Ricardo Alonso-Zaldivar and Jack Gillum. Government health care website quietly sharing personal data. https://apnews.com/article/31490a20926d4ed3b98ff2d0ed8fc81d, January 2015. Accessed on 2021-10-09.

[5] Inc. Brave Software. https://brave.com/.

[6] MDN Contributors. Network monitor. https://developer.mozilla.org/en-US/docs/Tools/Network_Monitor. Accessed on 2021-12-10.

[7] MDN Contributors. Referrer policy. https://developer.mozilla.org/en-US/docs/Web/HTTP/Headers/Referrer-Policy. Accessed on 2021-12-10.

[8] DuckDuckGo. Duckduckgo tracker radar exposes hidden tracking. https://spreadprivacy.com/duckduckgo-tracker-radar/. Accessed on 2021-12-10.

[9] Steven Englehardt and Arvind Narayanan. Online tracking: A 1-million-site measurement and analysis. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, page 1388–1401, New York, NY, USA, 2016. Association for Computing Machinery.

[10] Mozilla Foundation and Mozilla Corporation its contributors. https://www.mozilla.org/en-US/firefox/new/.

[11] Apple Inc. https://www.apple.com/safari/.

[12] The jQuery Team John Resig. What is jquery?. https://jquery.com/. Accessed on 2021-12-10.

[13] The StackPath Team Lance Crosby. Build your edge. https://www.stackpath.com/. Accessed on 2021-12-10.

[14] Timothy Libert. Privacy implications of health information seeking on the web. *Communications of the ACM*, 58(3):68–77, 2015.

[15] Timothy Libert. An automated approach to auditing disclosure of third-party data collection in website privacy policies. In *Proceedings of the 2018 World Wide Web Conference*, pages 207–216, 2018.

[16] Jonathan R. Mayer and John C. Mitchell. Third-party web tracking: Policy and technology. In *2012 IEEE Symposium on Security and Privacy*, pages 413–427, 2012.

[17] Joshua D. Niforatos, Alexander R. Zheutlin, and Jeremy B. Sussman. Prevalence of Third-Party Data Tracking by US Hospital Websites. *JAMA Network Open*, 4(9):e2126121–e2126121, 09 2021.

[18] Opera. https://www.opera.com/browsers.

[19] Dmitriy Akulov Prospect One. A free cdn for open source. https://www.jsdelivr.com/. Accessed on 2021-12-10.

[20] Cooper Quintin. Healthcare.gov sends personal data to dozens of tracking websites. https://www.eff.org/deeplinks/2015/01/healthcare.gov-sends-personal-data, January 2015. Accessed on 2021-10-09.

[21] Matt Riley and the Swiftype Team Quin Hoxie. A powerful search experience for your website-without the learning curve. https://swiftype.com/. Accessed on 2021-12-10.

[22] Font Awesome Team. Get vector icons and social logos on your website with font awesome, the web's most popular icon set and toolkit. https://fontawesome.com/. Accessed on 2021-12-10.

[23] The Blueconic Team. Unified, actionable customer data. https://www.blueconic.com/. Accessed on 2021-12-10.

[24] The Cybot Team. Cybot - eprivacy - today. https://www.cybot.com/. Accessed on 2021-12-10.

[25] The Financial Times Team. Upgrade the web. automatically. https://polyfill.io/v3/. Accessed on 2021-12-10.

[26] The Hotjar Team. Understand how users behave on your site, what they need, and how they feel, fast. https://www.hotjar.com/. Accessed on 2021-12-10.

[27] The Matomo Analytics Team. Google analytics alternative that protects your data and your customers' privacy. https://matomo.org/. Accessed on 2021-12-10.

[28] The Prismic Team. Make your website editable for the whole team. https://prismic.io/. Accessed on 2021-12-10.

[29] The Prospect One Team. Global lead generation network. https://www.prospect.one/eng/. Accessed on 2021-12-10.

[30] The ReadSpeaker Holding Team. Lifelike text to speech for your customers. https://www.readspeaker.com/. Accessed on 2021-12-10.

[31] The Relay42 Team. Turn data into meaningful relationships. https://relay42.com/. Accessed on 2021-12-10.

[32] The Retyp LLC Team. Convert and monetize your website traffic. https://optinmonster.com/. Accessed on 2021-12-10.

[33] The Tweakwise Team. Tweakwise, smart software. https://www.tweakwise.com/. Accessed on 2021-12-10.

[34] The Usabilla B.V. Team. Optimize your digital channels with customer feedback. https://usabilla.com/. Accessed on 2021-12-10.

[35] The Wingify Team. Fast growing companies use vwo for their a/b testing. https://vwo.com/. Accessed on 2021-12-10.

[36] The Zebrafish Labs Team. Powerful image processing, simple apis. https://imgix.com/. Accessed on 2021-12-10.

[37] The Zendesk Team. Live chat software for the best customer conversations. https://www.zendesk.com/service/messaging/live-chat/?from=zp. Accessed on 2021-12-10.

[38] Inc Team The Cloudfare. The easiest way to accelerate and secure your website. https://www.cloudflare.com/. Accessed on 2021-12-10.

[39] Inc Team The New Relic. Monitor, debug, and improve your entire stack. https://newrelic.com/. Accessed on 2021-12-10.

[40] Tikno Tikno. Exploring the acceptance for pixel technology implementation in facebook ads among advertisers in indonesia. *KnE Social Sciences*, 3(1):96–107, Jan. 2018.

Appendix A

Appendix

A.1 Website List

1. Governmental
   - https://www.rivm.nl/en
   - https://www.rijksoverheid.nl/onderwerpen/themas/familie-zorg-en-gezondheid
2. Health Insurance Companies
   - https://www.zilverenkruis.nl/consumenten
   - https://www.interpolis.nl/
   - https://zorgverzekeringhema.nl/
   - https://www.salland.nl/
   - https://www.unitedconsumers.com/
   - https://www.besured.nl/
   - https://www.promovendum.nl/
   - https://www.prolife.nl/
   - https://www.ohra.nl/
   - https://www.fbto.nl/
   - https://www.nn.nl/Particulier.htm
   - https://www.zorgdirect.nl/
   - https://www.onvz.nl/
   - https://www.ditzo.nl/zorgverzekering
   - https://www.just.nl/
   - https://www.jaaah.nl/
   - https://www.independer.nl/
   - https://www.vgz.nl/
   - https://www.zorgenzekerheid.nl/
   - https://www.zorgkiezer.nl/
   - https://www.defriesland.nl/
   - https://www.bewuzt.nl/
   - https://www.nationalacademic.nl/
   - https://www.promovendum.nl/
   - https://www.umczorgverzekering.nl/
   - https://www.unive.nl/
   - https://www.zekur.nl/
   - https://www.cz.nl/
   - https://www.anderzorg.nl/
   - https://www.menzis.nl/
   - https://www.dsw.nl/Consumenten
   - https://www3.aevitae.com/zorgverzekeringen/
   - https://www.iza.nl/
3. "Top-Clinical" Hospitals
   - https://www.cwz.nl/
   - https://www.asz.nl/
   - https://www.amphia.nl/
   - https://www.catharinaziekenhuis.nl/
   - https://www.dz.nl/
   - https://www.etz.nl/
   - https://www.franciscus.nl/
   - https://www.gelreziekenhuizen.nl/home/
   - https://www.haaglandenmc.nl/
   - https://www.hagaziekenhuis.nl/home/
   - https://www.isala.nl/
   - https://www.jeroenboschziekenhuis.nl/
   - https://www.maasstadziekenhuis.nl/
   - https://www.martiniziekenhuis.nl/
   - https://www.mmc.nl/
   - https://www.meandermc.nl/patientenportaal/home
   - https://www.mcl.nl/informatie-rondom-coronavirus

- https://www.mst.nl/
- https://www.nwz.nl/
- https://www.olvg.nl/
- https://reinierdegraaf.nl/
- https://www.rijnstate.nl/
- https://spaarnegasthuis.nl/
- https://www.antoniusziekenhuis.nl/
- https://www.viecuri.nl/
- https://www.zuyderland.nl/

4. University Hospitals

- https://www.vumc.nl/
- https://www.amc.nl/web/home.html
- https://www.umcg.nl/
- https://www.umcutrecht.nl/nl
- https://www.lumc.nl/
- https://www.mumc.nl/
- https://www.erasmusmc.nl/
- https://www.radboudumc.nl/patientenzorg

5. Online Pharmacies

- https://www.apotheek.nl/
- https://www.efarma.nl/index.html
- https://www.deonlinedrogist.nl/
- https://www.drogisterij.net/
- https://www.onlinedrogisterij.nl/
- https://www.plein.nl/drogisterij/
- https://bik-bik.com/
- https://www.medicijnen.nl/

6. General Practitioners

- https://www.hapwaalfront.nl/en/homepage-new/
- https://www.zorggroep-almere.nl/
- https://vanbrachtendorgelo.nl/
- https://huisartsenkopvanzuid.nl/
- https://mchooglede.saas-zorgcentra.yardinternet.nl/zorgverlener/huisartspraktijk-de-wit/
- https://www.gchillesluis.nl/
- https://huisartsenpraktijkmaasbracht.nl/
- https://www.huisarts-groningen-centrum.nl/
- https://www.huisartsengroepdeschans.nl/
- https://www.hhhpraktijk.nl/
- https://huisartslofvers.uwartsonline.nl/
- https://www.huisartspraktijkdeestafette.nl/
- https://www.medischcentrummariaparochie.nl/
- https://www.aarveldmedischcentrum.nl/
- https://huisartsenpraktijktuindorp.praktijkinfo.nl/
- https://huisartsenpraktijkveur.praktijkinfo.nl/
- https://mesritz.onzehuisartsen.nl/
- https://huisartsenpraktijkbeatrixplantsoen.uwartsonline.nl/
- https://www.huisartsenpraktijkverhaaren.nl/
- https://leijnsehuisarts.nl/
- https://rozet.onzehuisartsen.nl/
- https://hapommeren.praktijkinfo.nl/
- https://huisarts.praktijkinfo.nl/
- https://www.huisartsenpraktijkpereira.nl/
- https://www.borrendamme.nl/#/home
- https://www.praktijklinthorst.nl/
- https://www.medisena.nl/
- https://www.praktijktheeuwen.nl/
- https://www.huisartsenpraktijkdegreef.nl/Default.asp?&HTTPSHASH=
- https://www.icone.nl/
- https://huisartsenpraktijksterkendejong.praktijkinfo.nl/
- https://www.huisartsgroningen.nl/
- https://huisarts-zaandam.praktijkinfo.nl/
- https://www.huisartsenvoorburg.nl/
- https://www.huisartspraktijkbrouwersgracht.nl/
- https://vanderidder.praktijkinfo.nl/
- https://huisartsenleonardspringer.praktijkinfo.nl/
- https://huisartsdevries.nl/
- https://www.mcbz.nl/
- https://huisartsenpraktijkkorvel.praktijkinfo.nl/
- https://huisartsreedijk.praktijkinfo.nl/
- https://www.hellendaal.nl/
- https://doktersebrechts.nl/
- https://www.leyendak.nl/
- https://huisartsenpraktijkassendelft.praktijkinfo.nl/
- https://willems.praktijkinfo.nl/
- https://hpboudesteyn.praktijkinfo.nl/
- https://www.medi-mere.com/home
- https://hapdewit.uwartsonline.nl/Default.asp?&HTTPSHASH=
- https://huisartsenpraktijk-crop.nl/
- https://www.huisarts-vijfhuizen.nl/
- https://hacuserk.nl/
- https://blokker.praktijkinfo.nl/
- https://www.huisartspraktijkjoosen.nl/
- https://www.zonboog.nl/gcommoord/
- https://www.huisartsenpraktijkcityplaza.nl/
- https://www.huisartsenpraktijknoor.nl/

- https://www.huisartsspijkerman.nl/
- https://www.aaaeco.nl/Airco
- https://diemenzuid.gazo.nl/
- https://huisarts-kappeyne.nl/
- https://www.huisartsenpraktijk-matenpark.nl/
- https://www.medi-mere.com/home
- https://www.huisartsenpraktijkjberkhoudt.nl/
- https://www.gezondheidscentrumlelystaete.nl/
- https://mercuur.praktijkinfo.nl/
- https://www.huisartspraktijkternaard.nl/
- http://www.huisartstopouria.nl/
- https://huisartsenpraktijkhiemstra.uwartsonline.nl/Default.asp?&HTTPSHASH=
- https://stelder-huisartszaandam.praktijkinfo.nl/

7. Physiotherapists

- http://fysiotherapie4u.nl/
- https://www.fysio-en-manueletherapieamerongen.nl/
- https://www.fysioflerninghof.nl/
- https://www.fisicafysiotherapie.nl/
- https://fysiotherapiehoekvanholland.nl/
- https://www.fysio-dubbeldam.nl/
- https://www.spinesport.nl/fysiotherapie-maastricht
- https://www.fysiomasters.nl/
- https://fysiotherapievanbreestraat.nl/

8. Dentists

- https://www.dentalcare-oost.nl/
- https://aemstelgroep.nl/
- https://www.tandartszalmplaat.nl/
- https://tandartspraktijkdeboemerang.nl/
- https://www.dentalclinichagens.com/?geo=netherlands
- https://www.tandheelkundegoudsesingel.nl/
- https://daniels-tandheelkunde.nl/
- http://www.tandartspraktijkdavidson.nl/
- https://tpliefdepassie.nl/
- https://endopraktijk.nl/
- http://www.tpvanderwielen.nl/
- https://tpvlaardingen.nl/
- https://dentville.nl/
- https://mondzorgvoorkids.nl/

9. Health Institutes

- https://www.trimbos.nl/english/?about-us/trimbos-international-ti-
- https://www.avl.nl/en/

10. Other

- https://www.thuisarts.nl/
- https://izz.nl/
- https://www.dokteronline.com/en/
- https://www.alphega-apotheek.nl/
- https://lvvp.info/
- https://www.acibademimc.com/