

# Exploring AdTech

## and visualizing data flows on Austrian news sites

DIPLOMARBEIT

zur Erlangung des akademischen Grades

**Diplom-Ingenieurin**

im Rahmen des Studiums

**Medieninformatik**

eingereicht von

**Carina Pratsch**

Matrikelnummer 1228399

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Peter Purgathofer

Wien, 24. März 2020

---

Carina Pratsch

---

Peter Purgathofer



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.  
The approved original version of this thesis is available in print at TU Wien Bibliothek.

# Exploring AdTech

## and visualizing data flows on Austrian news sites

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

**Diplom-Ingenieurin**

in

**Media and Human-Centered Computing**

by

**Carina Pratsch**

Registration Number 1228399

to the Faculty of Informatics

at the TU Wien

Advisor: Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Peter Purgathofer

Vienna, 24<sup>th</sup> March, 2020

---

Carina Pratsch

---

Peter Purgathofer



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.  
The approved original version of this thesis is available in print at TU Wien Bibliothek.

# Erklärung zur Verfassung der Arbeit

Carina Pratsch

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 24. März 2020

---

Carina Pratsch



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.  
The approved original version of this thesis is available in print at TU Wien Bibliothek.

# Acknowledgements

At this point I would like to thank my advisor Peter Purgathofer, for introducing me to this topic and the support during this work, who always had time to talk and answer questions, for offering tea and sending interesting articles about the topic.

Furthermore, I would like to thank all people involved for their time and feedback to improve this work, as well as my family and friends for their support.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.  
The approved original version of this thesis is available in print at TU Wien Bibliothek.



# Kurzfassung

Wir werden jeden Tag mit Werbungen konfrontiert, wobei offline Werbung nicht persönlich ist und eine breite Masse von Menschen ansprechen soll. Online sind wir jedoch mit Werbung konfrontiert, die oft gezielt auf uns abgestimmt ist und auf gesammelten Daten über unser Verhalten und unseren Interessen basiert. Das Ziel dieser Arbeit war es, mit einer explorativen Auseinandersetzung Werbetechnologien (AdTech) zu erforschen und wie sie BenutzerInnen verfolgen oder beeinflussen können. Dafür wurden verschiedene Aspekte von AdTech untersucht und es wurde versucht einen Einblick in das komplexe Werbe-Ökosystem zu geben, um den BenutzerInnen die „relevanteste“ Werbung anzuzeigen, sowie wie Werbung dafür verwendet werden kann, um politische Entscheidungen zu beeinflussen, indem WählerInnen personalisierte Werbungen angezeigt werden, oft auch in Verbindung mit Fake News. Die *Datenschutz-Grundverordnung* (DSGVO) schützt die Privatsphäre von EU-BürgerInnen online und der Verarbeitung personenbezogener Daten muss explizit zugestimmt werden. Jedoch wurde gezeigt, dass die BenutzerInnen hier oft noch vor Schwierigkeiten stehen, die Zustimmung zu verweigern oder sie dazu gedrängt werden der Sammlung und Analyse ihrer Daten zuzustimmen. Während der Untersuchung von AdTech sind Nachrichtenseiten öfters durch ihre hohe Anzahl von Cookies und Trackern aufgefallen. Da sie für viele BenutzerInnen eine tägliche Informationsquelle sind, wurden 10 österreichische Nachrichtenseiten ausgewählt und analysiert, um das Bewusstsein über das Thema zu erhöhen. Dazu wurden quantitative Daten über Verbindungen zu Dritten und mit welchen Elementen die BenutzerInnen beim Besuch der Seite in Kontakt kommen, beispielsweise Cookies, Skripten oder Tracking-Pixel, gesammelt. Es wurde gezeigt, dass beim Besuch von 9 der 10 untersuchten Nachrichtenseite, ohne Interaktion der BenutzerInnen, bereits Cookies von Dritten gesetzt wurden und der Besuch oft nicht datenschutzfreundlich war. Um den BenutzerInnen einen besseren Einblick in die gesammelten Daten zu ermöglichen wurde eine interaktive Visualisierung implementiert und mithilfe einer Expertendiskussion und Benutzerinterviews evaluiert. Das Feedback zeigte, dass die Visualisierung das Potenzial hat, mehr Aufmerksamkeit auf die versteckten Datenflüsse zu lenken und das Bewusstsein für die online Privatsphäre zu erhöhen.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.  
The approved original version of this thesis is available in print at TU Wien Bibliothek.

# Abstract

Every day we are confronted with advertisements, whereby offline advertising is not personal and is meant to appeal to a broad mass of people. However, online we are confronted with targeted ads based on collected data about our behaviour and interests. The aim of this thesis was to explore advertising technologies (adTech) and how they may target or influence users through exploratory research. Different aspects of adTech were explored and an attempt was made to provide an insight into the complex advertising ecosystem to show the user the “most relevant” advertisement and how advertising can be used for political influences by targeting voters with personalized ads, often in combination with fake news. The *General Data Protection Regulation* (GDPR) protects the privacy of EU citizens online and the procession of personal data requires explicit consent. However, it was shown, that users often face challenges to deny consent or are nudged to accept the collection and analysis of their personal data. During the exploration of adTech news sites repeatedly attracted attention with their large number of cookies and trackers. As a daily source of information 10 Austrian news sites were selected and analysed in order to raise more awareness about the topic. Quantitative data was collected about third party connections and with which elements the user came into contact, for example cookies, scripts or tracking pixel. It was shown, that when visiting 9 of the 10 observed news sites with no interaction, third party cookies were set and the visit was not privacy friendly for the user. In order to gain better insights into the collected data an interactive visualization was implemented and evaluated with an expert discussion and user interviews. The feedback indicates the potential to raise more attention to the hidden data flows and to increase the awareness about online privacy.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.  
The approved original version of this thesis is available in print at TU Wien Bibliothek.

# Contents

<b>Kurzfassung</b>	<b>ix</b>
<b>Abstract</b>	<b>xi</b>
<b>Contents</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation and Problem Definition . . . . .	1
1.2 Aim of the Work . . . . .	2
1.3 Structure . . . . .	2
<b>2 Advertising Ecosystem</b>	<b>5</b>
2.1 Roles . . . . .	5
2.2 Tracking Methods . . . . .	7
2.3 Targeting Methods . . . . .	11
<b>3 Exploring AdTech</b>	<b>15</b>
3.1 Political Influence . . . . .	15
3.2 Ad Library . . . . .	22
3.3 GDPR and Privacy Notifications . . . . .	26
3.4 Dark Patterns and Nudging . . . . .	32
3.5 Tools to Visualize and Block Tracking Methods . . . . .	36
3.6 Discussion . . . . .	42
<b>4 Analysis of News Sites</b>	<b>45</b>
4.1 Collecting the Data . . . . .	45
4.2 Classification . . . . .	50
4.3 Other Findings . . . . .	54
<b>5 Interactive Design</b>	<b>59</b>
5.1 Design Prototypes . . . . .	59
5.2 Implementation . . . . .	65
5.3 Colours . . . . .	67
5.4 Labels . . . . .	68
	<b>xiii</b>

5.5	Elements . . . . .	70
5.6	Interactions . . . . .	72
5.7	Evaluation . . . . .	75
<b>6</b>	<b>Conclusion</b>	<b>81</b>
6.1	Discussion . . . . .	81
6.2	Summary . . . . .	85
6.3	Future Work . . . . .	86
	<b>List of Figures</b>	<b>87</b>
	<b>List of Tables</b>	<b>91</b>
	<b>Appendix</b>	<b>93</b>
	Classified Element Types . . . . .	93
	Top Domains per Case . . . . .	99
	<b>Bibliography</b>	<b>103</b>

# Introduction

## 1.1 Motivation and Problem Definition

In my bachelor thesis I investigated *Facebook's* privacy settings and as part of it a new profile of a 20-year-old man who lives in Vienna was created. The first advertisement that appeared was an invitation to join a certain politician. *Facebook* uses targeted advertising based on given or tracked information to show certain advertisements only to a specific audience, in this case a typical, potential voter. But how does this work exactly and may possible influence the user's behaviour?

Advertisement is part of our daily life, we hear them on the radio or see them for example in newspapers, on billboards or television. The offline advertisement is not personal, because it was created for a broad mass of people. However, online we are confronted with a lot of digital advertising based on data collection, tracking and the use of personal information to address the users individually. In connection with social media advertising becomes even more personalized as it is linked to personal information, for example age, gender, education, job, place of residence, friends or interests.

Advertising Technologies, short adTech, describes different types of techniques used by advertising companies to manage and distribute advertisements on different platforms. A main part of this industry is to collect personal information to display the "most relevant" advertisement to a targeted online audience. However, personalized or tracking based advertising, can not only be used to display ads for a specific product, that the user has recently viewed, it can also use the collected information for political targeting, to show specific advertisement in order to possibly influence the population or steer in a certain direction.

An example of this is the Brexit referendum in 2016, in which political advertising on *Facebook* spread fake news. It was investigated by the British Parliament [1] and they proved the use of dark ads or microtargeting. This is advertising that is only visible

to a specific target group based on personal data and therefore, it was possible that a party used contradictory statements which could not be discussed publicly because not everyone was aware of them.

On one side is the adTech industry, with the goal to increase their profits, on the other side are the users who are followed to collect and analyse personal data to show them advertisements that might interest them. This intrusion into one's privacy may not be known to all users.

### 1.2 Aim of the Work

The aim of this work is to explore adTech and the sphere of data collection and how it may target or influence users through exploratory research. AdTech has many players and potential areas of influence, and an attempt to give an insight into this ecosystem will be made. Depending on the findings a way to enlighten the users on the topic and encourage awareness of the problems of targeted methods leading to the intrusion of privacy will be determined and implemented.

### 1.3 Structure

The next chapter *Advertising Ecosystem* will describe the different roles involved in showing advertisements, how personal information is obtained with different tracking methods and which methods are used to target the user, through literature review.

The chapter *Exploring AdTech* will take on an explorative approach to look at different perspectives the user can come in contact with adTech to find current problems. First examples of the use of advertising technologies in political campaigns and elections are shown, as well as examples of targeted advertisements used to spread fake news and influence people to make a certain decision. The next section focuses on the *General Data Protection Regulation* (GDPR) of the EU and its realisation in the form of privacy notifications, asking the user for consent and leading to nudging and the usage of dark patterns. The last section tests different tools already available to visualize online tracking and methods to block adTech companies from tracking and collecting data.

In the chapter *Analysis of News Sites* quantitative data of ten Austrian news sites are collected, about how many connections to third party companies are established and which elements are loaded, which were classified into different types. The data is collected for different cases, to explore if the news sites are respecting the privacy of the user and what effects the acceptance of the default options of the privacy notifications has for the user.

The following chapter *Interactive Design* will focus on finding a form of visualization to gain insights of the collected data, and describes the design process and the implementation, leading to the evaluation of an expert to find improvements and interviewing users



to determine if the goal of the visualization, to encourage awareness on the topic, can be achieved.

In the last chapter *Conclusion* the insights gained are discussed and the findings of the thesis summarized. The last section ends with further improvements and future work.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.  
The approved original version of this thesis is available in print at TU Wien Bibliothek.

# Advertising Ecosystem

## 2.1 Roles

In the beginning of online advertisements, the ads were served directly by the website, were not personal and the user knew where they came from [2, 3]. The modern advertising ecosystem is becoming more complex in its goal to follow a user around the internet to collect as much personal information as possible. The user's privacy is traded for displaying targeted advertisements to increase the profit of this billion-dollar business [2, 4, 5].

The three main roles of the advertising ecosystem are the advertiser, the publisher and the connecting component the ad platform. The users are not considered directly part of this infrastructure as their data is only a tool and they receive no profit [2]. A more complex schema of the interactions between the players is shown in Figure 2.1 and these roles are described in the following.

### Advertiser

Advertisers want to show their advertisements about a brand or product to potential customers, based on demographics or market segments. They are looking for space on websites to display ads and are willing to pay that a targeted audience sees their advertisements or campaigns. The potential audience is on different publisher sites all over the internet, making it difficult to reach them. [2, 6]

### Publisher

A publisher is usually a website that produces or provides online content, for example a search engine, a news site or a blog. They earn revenue to provided space to display ads to the users and through the publisher the user comes in contact with adTech [2, 6]. From the publisher's perspective, "*online advertising is the pillar that sustains the Internet's 'free' content and services*" [2].

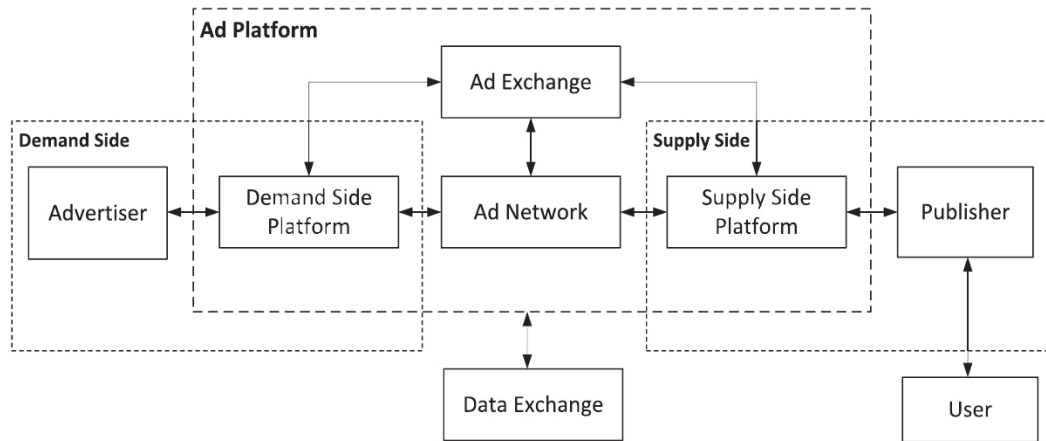


Figure 2.1: The interactions between the players in the adTech ecosystem, from [2].

### Ad Platform

Ad platforms bring the demand and the supply side together and contain different players that perform a particular roll to optimize the process. They match the provided ads with the specified targeted audience and the *“accuracy of said matching clearly depends on the ad platforms’ ability to track and profile users based on the information that can be mined from their online activity”* [2].

### Ad Network

Ad networks buy and manage the space for displaying ads from publishers, and sell to advertisers by different models. The *Cost per Mille*, allows advertisers to buy blocks of 1.000 views of the advertisement or the *Cost per Click*, where the advertisers pays each time a user clicks on the ad [6].

### Ad Exchange

These platforms use a modern form to supply advertising space thought auction. In *Real-Time Bidding* (RTB) advertisers bid in real time to show their ad to a specific user in real-time. Ad exchanges offers the space in an auction and provides information about the user who will see the advertisement. The advertiser now bid on the impression, depending on their targeted audience, and the winner displays the ad to the user. [2, 6]

### Demand Side Platform

Demand side platforms provide an interface for advertisers to help choose the audience and the right media for displaying advertisements. They manage the interaction with the ad distribution and bid for the advertiser in auctions. [2, 6]

## Supply Side Platform

The supply side platforms provide an interface for publishers to help manage the inventory of ad space and provide the resources to find targeted advertisements to maximize the profits [2, 6].

## Data Exchange

To make the right targeting and bidding decisions, the players from ad platforms need personal information about the user. The data exchange is provided by companies called data aggregators or data brokers, that collect or buy information and specialize in data mining [2, 7]. The data points are collected from different online and offline sources, like social media, web history, purchase history or from public records and can include information about full name, age, gender, address, telephone number, email address, income, education or occupation [7]. Information can also be obtained by free games or apps on mobile phones and if the user is not careful and allow access to, for example, contacts, location or history, the information can be collected and processed. The data brokers combine all information and create audience segments which can be bought [7].

## 2.2 Tracking Methods

To display the “perfect” advertisement to users, the adTech industry needs data to know more about them. The following tracking methods provide an overview of the most common techniques, which can occur individually or in combination. In general, a distinction is made between first party and third party tracking.

One basis of tracking methods is that the user establishes a connection to the third party and if the third party has achieved this, information is exchanged. The third party receives the IP-address of the user which can be used to determine the approximate geographical location of the user and information about the website from which the connection was established. The third party receives further information about the user’s browser and operating system which can be used for different tracking methods.

### 2.2.1 First Party Tracking

First party describes a website that is deliberately accessed by the user, e.g. a publisher. With first party tracking, personal information that the user provides through the use of the site is collected and analysed. Since the information depends on the direct use of the user, this form of tracking is difficult to classify as malicious or to block [2], but the privacy risks depends on the audience size of the publisher.

When using a social media platform, the risk is particularly high because the user consciously shares personal information with the platform. For example, *Facebook* requires data such as first-name, last-name, gender and birthday already at registration. Through the use of the social media platform even more personal data is collected, such

as relationship status, religious and political views, education, location, friends, likes and interests [8] as well as the interaction with the content.

### 2.2.2 Third Party Tracking

Whereas a third party is an entity or domain to which an indirect connection is established, to which the user is not always aware of. That happens with “*content embedded in first party sites from which user information is also leaked to third parties*” [2].

For example, if the user visits *facebook.com* directly, a connection to a first party is established. If, however, a like-button from *Facebook* is integrated on another domain, which loads elements from *facebook.com*, *Facebook* is to be regarded as a third party in this case, since no direct connection is consciously established from the user.

### 2.2.3 Cookie

A HTTP cookie is a small file sent from a website and stored on the user’s hard drive. The file contains information about the user, e.g. preferences or login information, and is returned to the website each time the user visits to customize the users experience. First party cookies were originally designed to help users during online shopping in 1994, the usage of cookies have since changed and are viewed more often “*as an invasion of privacy due to the information collection without their consent*” [9]. Third party cookies are used to track users on different websites [5] and can track the user’s behaviour by collecting information about the interaction with the website, e.g. clicking on an ad, product or article. Each time data is loaded from the third party domain, it can access the cookie and the stored information for displaying a targeted advertisement.

### 2.2.4 Cookie Synchronization

To overcome the restriction that each cookie can only be accessed by the domain that created them cookie syncing [5, 10] or cookie matching [2, 6] was created by the ad industry. This tracking technique allows third parties to share user data in ad exchanges to match the different assigned IDs to the same user [2, 5, 10]. This technique is routinely performed in an RTB ad auction to combine information about a user to determine whether the user matches the targeted audience and depending on the match bids to display the user the advertisement [6].

In a recent study about cookie synchronization [5] it was found that within a year 97% of 850 users were at least once exposed to cookie synchronization and “*the average user receives around 1 synchronization per 68 HTTP requests, and gets up to 6.5 of their userIDs synced*”. The paper describes the tracking method as follows: The user visits *website1.com*, which loads a third party script from *tracker.com*. This allows *tracker.com* to set a cookie on the user’s browser with the ID *user123* for future identification. The user continues browsing and visits *website2.com*, which loads an image from *advertiser.com* who then sets a cookie identifying the user as *userABC*. *Website3.com*

contains elements from *tracker.com*, but not from *advertiser.com* and therefore the third party does not know that the user visited *website3.com*. Both adTech companies collaborate, thus *tracker.com* sent a redirect request to *advertiser.com*, e.g. simplified like *advertiser.com?syncID=user123&publisher=website3.com*. Now the browser calls *advertiser.com*, which can now access its own cookie, knowing *userABC* visited *website3.com* and is the same user *tracker.com* knows as *user123*. With this both companies can merge the collected user information in the background and invade the privacy of the user.

Another problem of cookie synchronization is between a publisher and an advertiser. Publishers with high valued content have limited impressions that can drive the value up in auctions. When information about the user is shared by the publisher, called *information leakage* [4], to provide information to bid on, the advertisers and ad exchange platforms have the ability to find the user on other cheaper publishers and thus reaching the same audience with targeted advertising at a lower cost. This reduces the cost of impressions on the premium publishers and leads to loss of profits, if not all publisher use cookie matching [4].

### 2.2.5 Flash Cookie and HTML5 LocalStorage

As more users became aware of the privacy risks of third party cookies, they started using browser add-ons to block them [11] or using the option to disable cookies that modern browsers provide [12]. Thus, the adTech industry needed a new way to track the users and match their online behaviour. So called *zombie cookies* [11] or *evercookies* [6] emerged, that are capable to restore HTTP cookies after the user deletes them [11]. This method is possible by using Flash Cookies or HTML5 local storage to store the information additionally, so it can be restored after deletion.

	HTTP Cookie	Flash Cookie	HTML5 LocalStorage
size limit	4 KB	100 KB by default	5 MB by default
expiration	expire by default	permanent by default	permanent by default
location	within the browser	outside the browser	within the browser
access	same browser	multiple browsers	same browser
difficulty to delete	low	high	high
usage level	remaining	declining	increasing

Table 2.1: Difference between HTTP Cookie, Flash Cookie and HTML5 LocalStorage [2, 13]

Flash Cookies and HTML5 local storage are more persistent than HTTP cookies and therefore more efficient in tracking users, see Table 2.1. While the usage of Flash cookies is declining, as it depends on the user using the Flash-plugin, the HTML5 local storage

tracking is increasing. HTML5 local storage offers more storage capacity, is browser independent, harder to delete and offers a more persistent tracking method. [2, 13]

### 2.2.6 Fingerprinting

After completely deleting all cookies and local storages the tracking information is lost and a new way to passively track and identify users was researched, the browser fingerprinting [11]. As the web developed different browsers emerged, not all supporting the latest architecture standard [14]. To avoid this problem HTTP included the *User-Agent request header* in which the browser includes their name, version and platform they are running to avoid limitations. Modern browser allow furthermore access to settings, like language, installed plugins, fonts, time zone or screen resolution, making the browser fingerprint more unique [12, 14].

The first research on this tracking technique was conducted by P. Eckersley in 2010, implementing a fingerprinting algorithm, in the *panoptlick*-project [15]. The study analysed 470.161 browsers and found that 83,6% of the fingerprints were unique, and if *Java* or *Flash* were enabled 94,2%.

An apparent disadvantage of the fingerprinting method is the changing of the browser environment, by upgrading or installing new elements, and thus changing the unique fingerprint. However, the *panoptlick*-project concluded that with a simple algorithm it was possible to match the updated browser fingerprints with a success rate of 99,1%. This enables a large number of users to be identified, and only those with identical configurations are not clearly distinguishable. [12, 15]

Another form of fingerprinting is the canvas fingerprint. With this method an HTML5 canvas element is drawn, sometimes invisible to the user. Depending on the browser and available resources the rendering varies and is analysed to possible reidentify the user [2].

With the method of fingerprinting users can be identified and tracked without depending on information stored within the browser. Browser fingerprinting can also be used to reidentify a user after deleting all cookies and respawn them [2].

### 2.2.7 Tracking Pixel

This method describes a simple technique, that loads an image the size of one pixel from a third party domain and is also called “*web bug*”, “*web beacon*”, “*invisible image*”, “*invisible pixel*” or “*pixel tag*” [16]. The pixel is integrated in the publisher’s website and loads automatically when the user visits the site. When the pixel is loaded, tracking information is passed on to the third party to collect the online behaviour of the user.

A research conducted in [16] about the usage of this method, investigated how many images loaded with the `<img>` tag were pixels and found that from the top 500 most visited websites in 2018, 31% included at least on pixel. It is concluded that this old technique is still in use as it lacks countermeasures or to provide a “*backup solution*” [16] if another tracking method is successfully blocked by the user.



## 2.3 Targeting Methods

After giving an overview how the adTech system works and describing possible tracking methods, this section focuses on the different targeting methods, to display advertisement to a specific audience.

*“While each type of the targeting methods customizes the displaying ads in different ways, the main goal is to improve the perceived relevance of the ads displayed.” [17]*

### 2.3.1 Microtargeting

Microtargeting can be defined as *“advanced psycho-geographic segmenting which is based on an algorithm determining a series of demographic and attitudinal traits to distinguish individuals for each targeted segment”* [8]. The audience is selected by several attributes that can be combined in different ways and allows the advertiser to display personalized messages or offers to the user. With the help of social media, in which users disclose a lot of personal data, this method has evolved in recent years and can also predict with high accuracy the reaction of a targeted audience to a message or ad [8].

Microtargeting allows the advertiser to put together an individual audience and for example, to promote a new product, show advertisements only to men and women between the ages of 18 and 25 years who enjoy sports or gaming, are single and live in Vienna.

### 2.3.2 Nanotargeting

This is an extreme example of microtargeting, and was accomplished by specifying the targeting attributes via *Facebook Ads*, that only a specific individual was shown the advertisement [8]. This method is only possible if a lot of personal information about a person is known, which in combination only applies to this one person.

For example, the target group is defined as, only women, between 25 and 30 years old, who went to school  $x$ , graduated university  $y$  and are now working for company  $z$ .

### 2.3.3 Retargeting

Retargeting specifically targets a user with advertisement about a product they previously viewed on a different website. This targeting method relies mostly on cookie matching and bidding in RTB auctions to retarget the user [6].

For example, the user is looking for a new smartphone and checks out a new model on a shopping site, but decides against buying it. Later, the user reads online news and is shown the exact same model that he or she has looked at before, besides an article.

### 2.3.4 Behavioural Targeting

This method focuses on targeting a user based on past behaviour. Previous collected information about the user's interests and personality are analysed to show personal advertisements [17, 18].

For example, a user is interested in computer games and a new game will be released soon that matches the collected interests, then this user is shown advertising for the new game. This method also specializes in predicting behaviour, for example, a user is interested in a specific location and may have also visited a page describing holidays destinations, then this user is targeted with advertisements about flights or hotels in that region.

### 2.3.5 Contextual Targeting

Contextual Targeting focus on providing advertisement based on the context the user is viewing. The effectiveness of this method relies on providing relevant ads for the shown content through keywords. Thus, resulting in a higher intent to purchase and a less intrusive user experience. [17, 18]

For example, the user reads an article about sleep disorders and is immediately shown the appropriate medication as advertisement beside the article.

### 2.3.6 Demographic Targeting

Demographic targeting shows advertisement based on location, gender and language to reach the audience [18].

For example, a global cosmetics line can only target women and displays different ads depending on location and skin, which appeals to the individual user and is in the appropriate language.

### 2.3.7 Geolocation Targeting

This method has “*one of the most dynamic targeting conditions*” [18] as users often change their location. In browsers the IP-address will provide insight of the general user location, whereas for mobile devices GPS is often used.

For example, the user is shown advertisements of restaurants, fitness centres or events in the near area. But the method can also be used to target users based on the average household income in their neighbourhood [18].

### 2.3.8 Political Targeting

This targeting method is particularly popular in the US during elections [8, 19] and is also becoming more and more popular in Europe [19]. Political targeting combines several targeting methods. First, the individual demographic information is important, then the behaviour of the person in the past and last using microtargeting to show the

voter content that appeals to them in particular, in order to convince to vote for the party.

Previously, political targeting was used offline and was not as detailed, but with modern methods of data collection and analysis it is now possible to individual address each voter and predict the response. Political parties buy this information from data brokers, such as *Cambridge Analytica*, who claimed to have collected information from about 230 mio. American voters, which includes up to 5.000 data points per person. Because of the privacy protection regulations in the EU “*online political microtargeting might not happen on a scale like in the US*” [19].



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.  
The approved original version of this thesis is available in print at TU Wien Bibliothek.

# Exploring AdTech

In the previous chapter an overview of the adTech ecosystem was given, the tracking methods explained and how they can possibly be used. This chapter will focus on exploring various examples to illustrate the impact and influence of adTech and why it is important to inform people about the topic. First, the British Brexit referendum and the American presidential election will be used to show how targeted advertising technologies are used to steer people to make a certain choice, even by spreading fake news. Then follows the EU's response through new regulation to prevent influence in elections and an example of the implementation via *Facebook's Ad Library*. The next section discusses how the GDPR is implemented in the form of privacy notifications, and shows examples of how the users are nudged to agree to tracking and the collection of personal data. The last section shows different tools already existing to visualize tracking methods, as well as tools that can be used by users to protect their online privacy.

## 3.1 Political Influence

### 3.1.1 Brexit Referendum

Two years after the Brexit referendum in 2016, in which Britain's voted to leave the European Union, the British Parliament investigated the spreading of fake news, which describes false or misleading content, and potential threats to democratic values [1]. A main part were the political campaigns on *Facebook* microtargeting the audience. These advertisements are also called *dark ads* or *dark post* and are only visible to the targeted audience. They were not publicly available and did not appear on the advertisers *Facebook* page [20]. *Facebook* was not very cooperative, but the personalized advertisements that were displayed to users were finally made available. This includes ads from the campaigns "*BeLeave*", "*50 Million*" and "*Vote to Leave*", with additional information about the targeted audience [21]. In the following examples of *dark post* are shown, which are intended to address different target groups in order to nudge them to leave the EU.

### 3. EXPLORING ADTECH

---



Figure 3.1: Leave EU: Sending £350 mio. a week to the EU, [22].



Figure 3.2: Leave EU: £350 mio. to the National Health Service (NHS) instead of the EU, [23].



Figure 3.3: Leave EU: £350 mio. to the EU or to Yorkshire, [23].

The main point of the Leave-campaign was the claim, that the United Kingdom could save £350 mio. a week, by leaving the European Union (see Figures 3.1, 3.2 and 3.3). This figure was obtained by dividing the official payment to the EU and has since proven to be misleading [24] as it leaves out a rebate and other EU fundings. The Treasury committee of the UK Parliament has calculated the amount to about £110 mio. [24]. The advertisements were targeting different audiences, Figure 3.1, was shown mostly to men (62%) [25], whereas the highest target group with 17% were between 45-54 years old. The second ad about health care, see Figure 3.2, addressed people over 45 years (85%) [26], focussing on men with 46% and women with 39%. The third ad about priorities was shown to a younger audience, targeting people between 18-34 years (64%) [26], whereas 30% were women and 34% were men. In 2018 a study found *“that 42 per cent of people who had heard of the claim still believe it is true, while just 36 per cent thought it was false and 22 per cent were unsure”* [27]. This shows that even after the fake news has been disproved it is not so easy to convince people and the false information stays in the public for a long time.

Other advertisements address the part of the British population that fear about immigration, in order to win over voters. An ad shows the high unemployment rates in other European countries compared to the UK and in the background, there is a man lying on the couch, watching TV (see Figure 3.4). The youth unemployment rate describes the percentage of 15-24 year-olds, *“who report that they are without work, that they are available for work and that they have taken active steps to find work”* [28] and is thus not the percentage of the jobless young people in the overall youth. The targeted audience was mainly men between 35-44 years (30%) and 45-54 years (23%) [26], outside the affected age group.

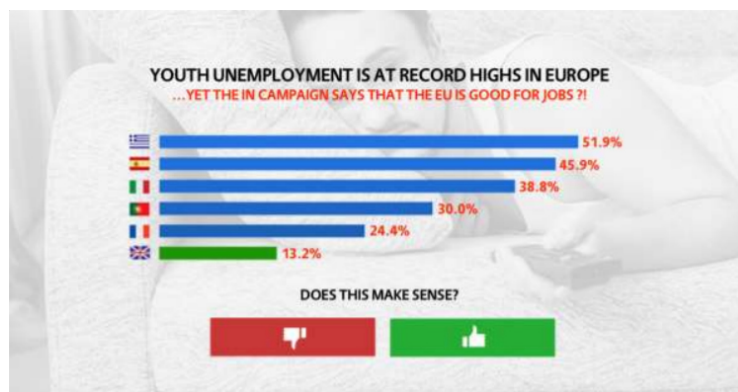


Figure 3.4: Leave EU: British youth unemployment compared to other EU countries, [23].

Another claim was that countries like Turkey, Albania, Macedonia, Montenegro and Serbia, who are still talking to the EU about accession, are joining. For each country there were advertisements showing the annual wage difference between the country and the UK as well as the number of inhabitants and asking *“good news?”*, an example can

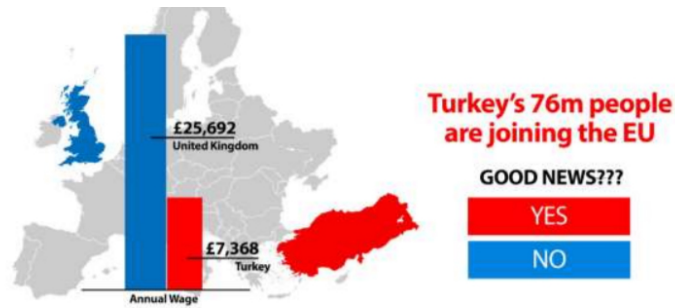


Figure 3.5: Leave EU: Turkey joining the EU, [23].



Figure 3.6: Leave EU: Visa-free travel for Turkish citizens and new border of the EU besides Syria and Iraq, [23].

be seen in Figure 3.5 showing the ad about Turkey. The audience of the dark ad was focussed on people between 35-54 years (76%) [26], targeted men and women equally and was subtitled with:

*“The next 5 countries joining the EU have poor populations with serious problems. The average annual income in Turkey is only £7.3k. Seriously. We already send £350 million to the EU every single week. Shouldn’t we spend our money on our priorities instead? Vote Leave Day - 23 June. Let Them Join? Click No! Vote Leave”* [26]

A further example, which claims that Turkish citizens can travel visa free to the EU, is shown in Figure 3.6, with the main target group of men between 55-64 years old (23%) [26]. The ad further emphasises how close the country is to Syria and Iraq, and the possible new border of the EU and thus of the UK. This statement was based on a deal that the EU made with Turkey in 2016 to stop the flow of refugees to Europe and allow visa-free travel if 72 requirements are met, which have not yet been fulfilled [29].



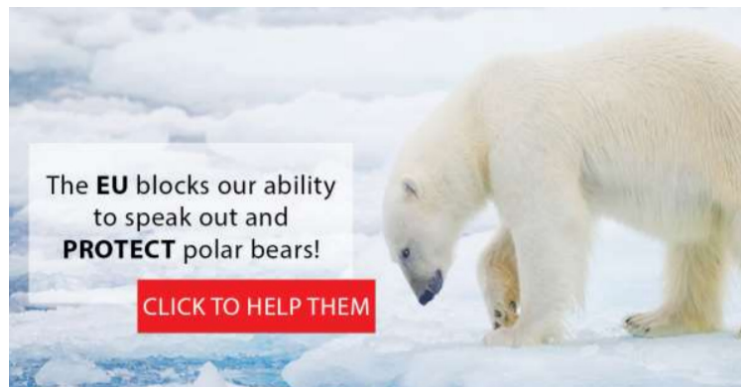


Figure 3.7: Leave EU: EU blocks ability to protect polar bears, [23].



Figure 3.8: Leave EU: EU wants to ban tea kettles, [23].

Other examples are ads which target women between 35-54 years (60%) [26] about animal welfare, which the EU prevents (see Figure 3.7), with different variants picturing polar bears, bulls, sheep or whales, or the claim that the EU wants to ban tea kettles which is targeted by region, with England (83%), Northern Ireland (4%), Scotland (7%) and Wales (6%) [26], which can be seen in Figure 3.8.

The last statement, which also had an ad about toasters, can be traced back to an EU plan to introduce new eco-design standards to make household appliances more energy efficient [30]. This was seen by the British population as a personal attack, as one third of all tea kettles in the EU are sold in the UK and the advertisement was subtitled with:

*“The EU plans to launch a kettle and toaster crackdown after the Brexit vote and has hidden it from the public fearing an announcement would push Britain towards leaving the EU. If we vote remain we will be powerless to prevent an avalanche of EU regulation that Brussels is delaying until after the referendum. Vote to leave the EU on 23 June. Save Our Cuppa. Vote Leave! Vote Leave”* [26]

As these examples show, the fake news is often based on a true statement, but has been twisted to misinterpret it. The aim was to generate outrage in the targeted audience in order to persuade the voters to make a certain decision, to leave the EU. The advertisements have a personal component, such as “*you decide*”, “*does this make sense?*” or “*good news?*” and offer the user a choice, for example “*EU*” or “*Yorkshire*”, “*thumbs up*” or “*thumb down*”, “*yes*” or “*no*”, to get the user to interact with the ad and click on it to get more information or fake news.

Another investigation of the British parliament was about the involvement of Russia in the referendum and the following elections. There is evidence of targeting techniques “*to amplify extreme voices in the campaign, particular those on sensitive topics such as race relations and immigration*” [31] and using misinformation as an “*unconventional warfare, using technology to disrupt, to magnify, and to distort*” [31]. The results of the latest investigation in 2019 about the topic is still withheld from the public and causes a loss of trust among the people in democratic values [32].

These example shows that fake news can remain for a very long time and influence democratic decisions. The problem with microtargeted advertisements is that they are only shown to a specific target audience. They appeal to them in a very personal, emotional way and sometimes prevent public discussions, since not all people are aware of them. Advertisements shown between other posts on social media and having the same format as other posts may not always be perceived as advertisements, misleading the audience. Another problem is the source of political advertisements, because anyone can place an ad from anywhere. Someone from Russia could have easily place an ad on *Facebook* and specified the demographic audience to the UK.

#### 3.1.2 US Elections

Targeted Advertisements, especially on social media, played a major role in the US presidential elections in 2016. The company *Cambridge Analytica* was involved and “*created psychographic classifications of voters by harvesting Facebook users posts, likes, and social networks and matching them with their comprehensive voter profile data*” [20]. Over 50 mio. *Facebook* profiles of US citizens were obtained without consent and *Facebook* denied the data breach [33].

The collected data allowed personal messages to appeal to a targeted audience. For example, could people interested in guns be targeted with “*Hillary will take away your guns*” while family-oriented people would receive “*guns protect your loved ones*” [20]. Donald Trump’s campaign used dark posts to target African-American voters to remind them that Hillary Clinton called African-American males “*super predators*” to influence the voting behaviour or suppress the voters [19]. This highlights a major problem of targeted advertising, that without sharing the displayed advertising, other people would not have known about it and may only wondered why fewer people voted from a specific group. If this happens shortly before the election, it is not always possible to react to it or to counteract potential fake news or influences.

In 2016, *Facebook* reached 67% of all American adults, whereas two-thirds used the social media platform as their news source, which corresponds to 44% of the population [34]. Personalized news feeds created an echo chamber and the more users clicked on and shared fake news, the more fake news was suggested to them. “*Before social media, the filter was provided by media companies, who acted as gatekeepers to the news and had staff trained in fact-checking and verifying information*” [35]. A study from 2016 researched online misinformation [36] and found, that it took about 10 to 20 hours to debunk fake news and share the fact-checked content. Furthermore, fake news are provided “*by few very active accounts, and grass-roots responses that spread fact checking information several hours later*” [36], highlighting the problem of the usage of fake news during the presidential election. This shows that it takes time to disprove misinformation, but then it may happen that the fact-check does not reach the respective people, because they are in their echo chamber.

*Facebook* was not constipated as a news business and cared more about the engagement with the content than the content itself. The company was further criticised that their team of editors censored content and suppressed conservative news in favour of republican content, which led to the editors being replaced by an algorithm, which in turn led to fake news ending up in the news feeds more often. Sharing the fake news also makes it more credible, because there is a trust between friends and they are more likely to believe the news and spread it themselves. [35].

After the election a foreign involvement was also investigated. *Facebook*, *Google* and *Twitter* testified that advertisements were purchased by Russian linked groups [19]. *Facebook* admitted that 3.000 advertisements can be linked to the same groups and that the ads were “*primarily focused on divisive social and political issues such as guns, LGBT rights, immigration, and race, and targeted specific categories of individuals*” [19].

In the current presidential elections, a debate about the influence of targeted political advertisement is still ongoing. *Facebook* is refusing to take down political advertisements even if they are proven to be misleading or fake [37]. Mark Zuckerberg, the CEO of *Facebook*, defends this decision as a commitment to free speech and does not want to censor people. As more people use social media platforms as an information source, fake news prevents people to make an uninfluenced choice, an important part of a democratic society.

### 3.1.3 EU Elections

In response to Russian influences in elections, advertising that displays fake news and the *Cambridge Analytica* data breach the European Commission proposed measures to counteract these influences in September 2018. The measures aim to provide more transparency in political advertising to protect democratic values for fair elections without foreign influences or manipulative advertisements, that stops people from making an informed choice [38].

Political parties were urged to apply the following measures, from [38]:

- EU citizens should easily recognise online paid political advertisements and communications, as well as the party, foundation or organisation behind them.
- Information about the spending for online activities should be made available, including paid online political advertisements and communications, as well as information about targeting criteria that are used to distribute such advertisements and communications.
- Paid online political advertisements and communications should be made accessible through the parties' websites.

Another point was the data protection of personal information and introducing new sanctions if the data was not appropriately protected or an attempt to influence elections was found. In March 2019 the new rules regarding online political advertisement in the EU were enforced and all advertisers now need to be authorized in their country to fight foreign abuse.

To comply to the rules *Facebook* added an *Ad Library* to view all political advertisements and individual ads are labelled with “*Paid for by*” at the top [39]. Advertisers now need to submit documents proofing their identification and location and information about a party, like campaign budget and how many people viewed the ad along with their demographic information. *Facebook* sees the implementation of this measure as “*Committing to Transparency and Accountability*” [39] and is working against identity theft.

*“We’re up against smart, creative and well-funded adversaries who change their tactics as we spot abuse. But we believe that they will help prevent future interference in elections on Facebook. And that is why they are so important.”*  
[39]

This shows *Facebook’s* different stance in the US and EU and that it needs appropriate regulation to compel the companies to make changes. Companies do not want to voluntarily admit data breaches, misuse of their platforms or lose a source of income through advertising that is particularly provocative and therefore often liked, shared and clicked on. Furthermore, it is not always clear to users what happens to their personal data or how it might be used against them.

## 3.2 Ad Library

In *Facebook’s Ad Library* information about all collected ads about “*politically or socially relevant topics*” [40] can be viewed depending on the available country. The recording

Seitenname	Disclaimer	Ausgebener Betrag	Anzahl der Werbeanzeigen in der Bibliothek
Pamela Rendi-Wagner	SPÖ	282.196 €	1.558
SPÖ	SPÖ	259.948 €	1.549
Norbert Hofer	FPÖ	218.939 €	189
Die Grünen	Die Grünen	182.989 €	327
Sebastian Kurz	Volkspartei	180.531 €	1.691
Harald Vilimsky	FPÖ	119.397 €	131
European Parliament	European Parliament	103.793 €	277
FPÖ	FPÖ	97.482 €	128
Greenpeace Österreich	Greenpeace Österreich	95.377 €	3.945
Dominik Nepp	FPÖ-Wien	77.526 €	221

Figure 3.9: Ad Library: *Facebook* pages with the highest ad spending in Austria.

started in March 2019 and can be filtered by time periods. For example, for Austria 69.716 related advertisements are available with a total amount of €4.286.814 spent between March 2019 and February 2020.

The information can be sorted in different ways or the user can search for a specific advertiser. In Figure 3.9 the data from March 2019 to February 2020 is sorted by *Facebook* pages with the highest spending, which are mostly political parties, whereas the page with the most recorded advertisements is *Greenpeace* with 3.945 ads.

The number of advertisements can also indicate whether they are possibly only placed for a certain target audience. As the period of one year is shown here, *Greenpeace*, for example, could also have only addressed many different topics during this period.

Before political elections it is interesting to look at these figures in a shorter time span. Since presidential elections are scheduled for November 2020 in the United States, the report also includes current political advertising election spendings. For example, the “*Mike Bloomberg*” page spent a total of \$1.219.840 on 20.183 ads on one day (February 24, 2020), which indicates that there are many targeted advertisements that address the same topic, but are formulated or designed differently depending on the target audience to address the targeted people as personally as possible.

### 3. EXPLORING ADTECH

Further information are displayed, about in which state the most spending was made, which sites are the top spenders in this state, as well as the total amount spent. If enough information is available, the most frequent search queries from the last week are also displayed, in order to find out which topics are currently relevant. Furthermore, it is possible to download a report of the selected period of time containing all the advertisers, their spending and the number of advertisements, total and sorted by state.

Upon selecting an advertiser more information is provided, such as when the page was created, which country is specified as the main resident of the operators, or the total spending on advertisements. All ads of the page are shown and can be sorted by displayed country, time period, whether they are still active or by platform. In Figure 3.10 two ads from *Greenpeace* can be seen, the first one is still in circulation and the second one is already inactive. As *Facebook* also owns *Instagram* information which platform displays the advertisement is included, as well as different known versions of the ad.

The figure displays two advertisement cards from the Ad Library. The first card is marked 'Aktiv' (Active) and shows a Greenpeace advertisement for 'Rette den Amazonas' (Save the Amazon). It includes the text: 'Im Amazonas werden pro Minute 1042 Bäume gefällt, darunter auch jahrhundertalte Baumriesen! Die Gier nach Fleisch, Tropenholz und Aluminium zerstört in Windeseile eine der letzten Wildnisse unserer Erde. Uns bleibt nicht mehr viel Zeit! Helft uns jetzt, den Amazonas vor den Kettensägen zu retten: Bitte unterzeichnet die Petition und leitet sie an eure Freunde weiter! Danke!' and features an image of a colorful drum in the rainforest. The second card is marked 'Inaktiv' (Inactive) and shows a Greenpeace advertisement for 'Stoppt die Bienenkiller' (Stop the Bee Killers). It includes the text: 'Seit Jahren sterben massenhaft Bienen! Viele von ihnen vergiften sich an den Bienenkiller-Pestiziden aus der industriellen Landwirtschaft. Und obwohl die europäische Behörde für Lebensmittelsicherheit längst Kriterien für ein bienenfreundliches Zulassungsverfahren entwickelt hat, werden diese einfach nicht eingesetzt! Zu groß ist die Lobby der Agrarkonzerne, die um ihre Gewinne fürchtet. Bitte hilf uns das Bienensterben zu stoppen...' and features an image of a bee in a field. Both cards include a 'Mehr dazu' (Learn more) button and a link to 'Anzeigendetails ansehen' (View ad details).

Figure 3.10: Ad Library: Two examples of advertisement from *Greenpeace*.

For each advertisement more information about the audience can be viewed, based on gender and age groups as well as the region the advertisement was displayed. Furthermore,

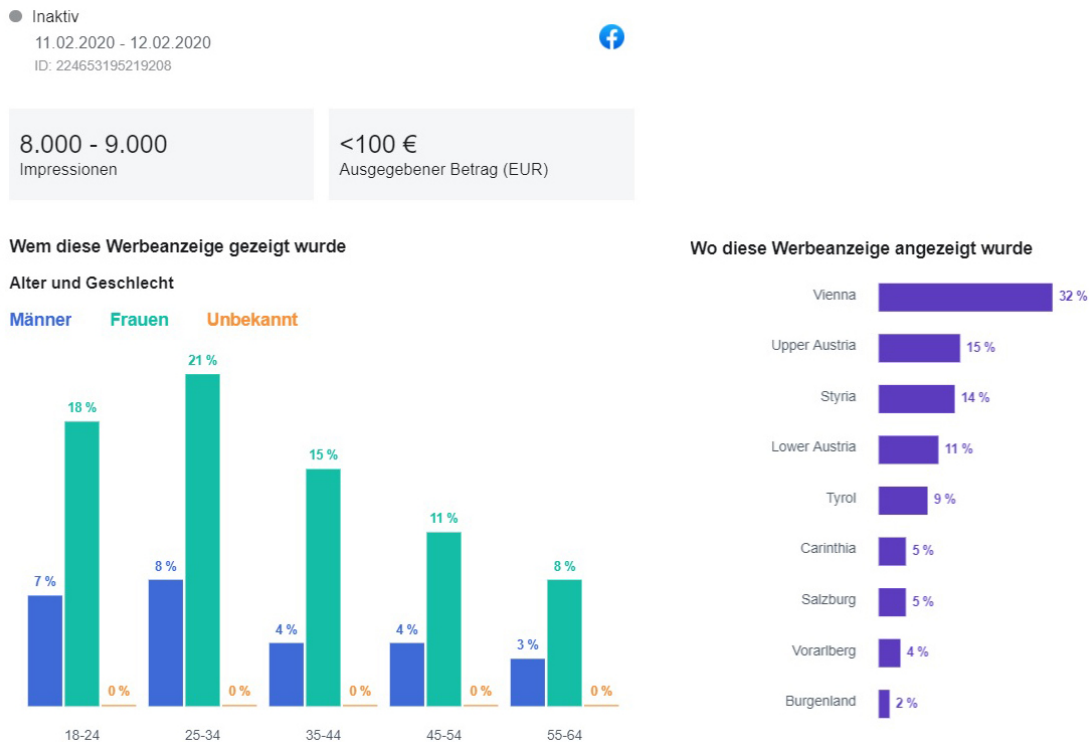


Figure 3.11: Ad Library: information about the ad as well as demographics of the targeted audience.

it is shown in which time period the ad was displayed, how much was spent on it and an estimate of how many people have seen the advertisement. The already inactive ad about dying bees (see Figure 3.10) was only active for two days on *Facebook*, but had about 8.000 to 9.000 impressions, as seen in Figure 3.11. It can also be seen that the advertisement was primary shown to women (73%) of different age groups, whereas the main group was between 25 and 34 years old (21%). Furthermore, it can be seen in which state the advertisement was shown to the targeted audience as well as the targeted value, in this case the most targeted group with 32% lived in Vienna.

This example shows that even with less than €100 it is possible to reach many people in a short period of time. But *Facebook* allows at the creation of an advertisement to select the audience also according to criteria such as: education, job, titles, interests, behaviour, device usage, or selecting the audience based on connections to *Facebook* pages or events [41]. It is also suggested to add the *Facebook* pixel on the website to track the behaviour of the users, to target them later individually. Furthermore, it is possible to create a custom audience based on information about customers collected otherwise, such as first name, last name, email address or contact information. All this information, how the targeted audience was exactly selected, is not visible in the *Ad Library*.

### 3.3 GDPR and Privacy Notifications

Personal data is collected from users through the use of websites. The *General Data Protection Regulation* (GDPR) of the EU state that the processing of personal data needs legitimate interest or the consent of the person.

*“Consent must be freely given, specific, informed and unambiguous. In order to obtain freely given consent, it must be given on a voluntary basis. The element ‘free’ implies a real choice by the data subject. Any element of inappropriate pressure or influence which could affect the outcome of that choice renders the consent invalid.”* [42]

Cookies are mentioned in the GDPR once in Recital 30 *“Online identifiers for profiling and identification”*, which mentions, that cookies and the IP-address are considered personal data if they can be used to identify the person [43].

*“Natural persons may be associated with online identifiers provided by their devices, applications, tools and protocols, such as internet protocol addresses, cookie identifiers or other identifiers such as radio frequency identification tags. This may leave traces which, in particular when combined with unique identifiers and other information received by the servers, may be used to create profiles of the natural persons and identify them.”* [44]

This led to consent or cookie notices shown to user when they visited a website the first time, asking to consent cookies and allowing third parties to collect and analyse personal information.

A recent study examined the impact of the GDPR on third party cookie usage and found *“that on average the number of third parties dropped by more than 10% after GDPR”* [45]. Another aspect of the study was to examine the browser data of users for a year. No impact was found, concluding that a lot of websites only offer to accept the usage of cookies and thus leaving the user only with the choice to accept or be excluded. In other cases users may not use the option to manage the settings and just accept the default options and thus accepting more third party cookies.

In October 2019 the Court of Justice of the European Union ruled to protect against personal data collection, *“that while consent could include ticking a box on a website, ‘silence, pre-ticked boxes or inactivity’ does not constitute consent”* [46]. This ruling should reduce the number of tracking cookies for the average user that accepts the default options in the future, but there are still problems regarding these “cookie banners”.

In the following a few examples are shown, sometimes misleading the user or making it very hard to opt-out of not giving consent. Cookies, that are strictly necessary do not require additional consent and the user usually cannot opt-out. Every publisher that



provides content to citizens of the EU has to comply to the EU regulations, but the adoption of the regulations takes time and are sometime only casually implemented.

A simple notice, stating the value of privacy can be seen in Figure 3.12, giving the user the choice to accept cookies or leave the website. The notice only states the use of cookies, but in the privacy policy the automatically collected information is also obtained by the use of *“tracking technologies such as cookies, web beacons, tracking pixels, and local shared objects, also known as flash cookies, to collect information relating to you and your use of the Service”* [47].

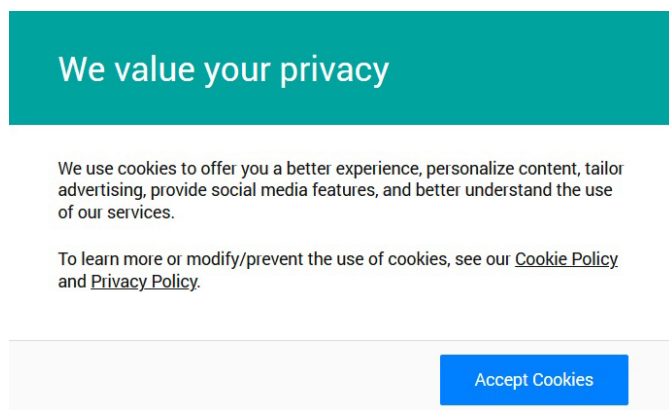


Figure 3.12: A simple cookie notice, stating how first party cookies are used, with one option to accept the usage

In Figure 3.13 the cookie notice accepts no clearly given consent, as the dismissal or to continue browsing is also interpreted as acceptance. This is misleading to some users, since they may assume that by closing the note, they have not accepted them and the GDPR states that *“consent must be freely given”* [42].

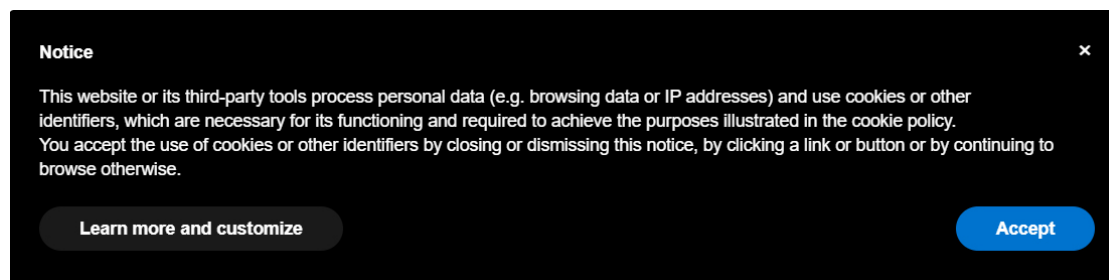


Figure 3.13: This privacy notification accepts closing, dismissal, clicking on links or buttons, continuing browsing and accepting as acceptance

To show that it is always important to look at the preferences and to whom you may consent the following example was found on a cooking website. The harmless looking

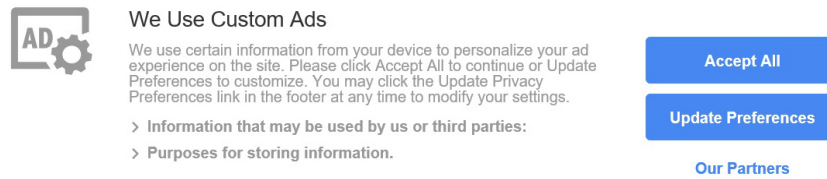


Figure 3.14: This simple cookie notice leads to advanced settings.

cookie notice (see Figure 3.14) proved to be time consuming in an attempt to view a recipe without giving consent that third parties collect information. First an advertisement blocked the navigation, leading to confusion about how to navigate in the cookie notice settings, as seen in Figure 3.15. The type of the ad could be targeted, do they know that I am a bad cook and should rather order frozen foods or contextual and some of their products appear in the article. It is targeted demographically because the ad is in German and the company delivers in my country. The advertisement raises the question whether personal information has not already been collected by third parties without explicit consent. Furthermore, the notice indicates that the use of the site is only possible by consenting to the use of personal data for the listed purposes.

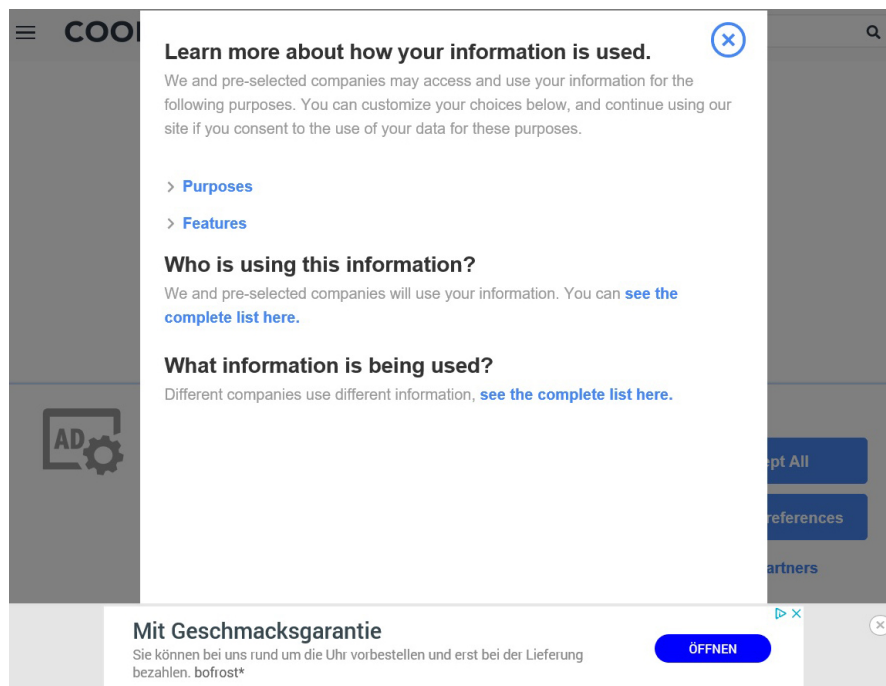


Figure 3.15: The overview of the advanced cookie settings blocked by advertisement.

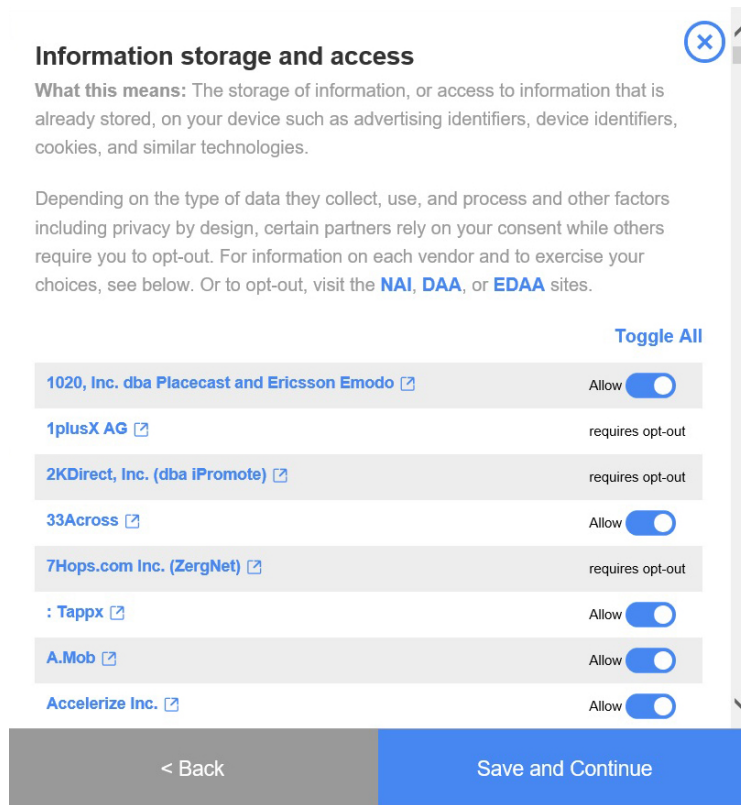


Figure 3.16: Third parties can be toggled to deny consent or require individual opt-out.

The cookie notice provided a list of all companies that will have access to the collected information, which added up to 536 unique companies. The options “*Purposes*” and “*Features*” lead to eight sub-options for which the preferences had to be set individually. An example of a sub setting “*Information storage and access*” (see Figure 3.16) shows that a toggle option to deny consent is available, but a lot of companies require a direct opt-out instead of on opt-in. In this subsection 505 companies were listed, 374 could be toggled while 131 required an opt-out, the highest opt-out value was 195 out of 425 companies in “*Measurements*”. Visiting the provided links to opt-out websites showed that one had 130 participating companies, leading to the conclusion that the user needs to opt-out on all three websites or some opt-outs are only possible by visiting the company directly. A few companies were randomly selected which lead to more cookie notices that needed to be managed or websites blocked by the browser due to security risks.

This example shows that the execution of the GDPR is not always user-friendly and the results from [45], that user often just accept the default options and simply click “*accept*”, can be understood, if they have to make these efforts just to read a recipe, without personalized advertisements and agreeing to the collection of personal data.

## No ads on this website! Just pageviews

This website protects your privacy by adhering to the European Union General Data Protection Regulation (GDPR)...but we'd also like to see how many people visit our website by viewing analytics & conversion data. This means NO ads on our website. Please state below which processes you consent to. We will not use your data for any other purposes.

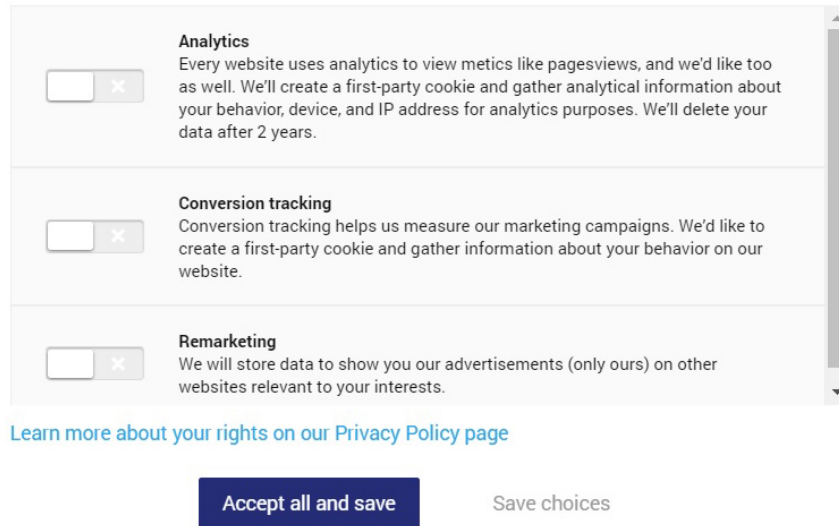


Figure 3.17: A cookie notice highlighting the “Accept all and save” button to nudge the user to accept and hiding “Save choices”.

The next two examples follow the regulations that no options are pre-selected and the user has the choice to deny not necessary options. But in Figure 3.17 the highlighted button “Accept all and save” tries to nudge the user on clicking it and thus may overlook the light grey option, that does not look like a button, of “Save choices” to deny consent. The privacy notification further indicates that although no advertisements will be displayed on this site, information will be collected to display targeted advertisements on other sites upon consent and the privacy policy states that with this consent *Google AdWords* will collect cookie-based behavioural data about the user. In Figure 3.18 the necessary cookies are selected, which is GDPR conform, but the user may be tempted to select the highlighted button “select all” believing to confirm the selection or non-selection. The “save” button appears to be disabled and the link for more details is hidden in light grey at the bottom. In the case of accepting all options, the user would also agree that, in addition to statistical information about the behaviour on the website, marketing cookies from third parties may be set to enable personalised advertising and by loading social media content, enabling them to track online behaviour.

In Figure 3.19 the “Analytics Cookies” are preselected and after opting-out a new button stating “Allow All” appears in place where previously the button “Save Settings” was

### Cookie-Einstellungen

Wir nutzen Cookies auf unserer Website. Einige von ihnen sind technisch notwendig, während andere uns helfen, diese Website zu verbessern oder zusätzliche Funktionalitäten zur Verfügung zu stellen.

Notwendige Cookies     Statistik  
 Marketing     Externe Medien

Alle auswählen

Speichern

Details anzeigen ▼

Impressum | Datenschutz

Figure 3.18: A cookie notice nudging the user to click “select all” to accept the usage of cookies.

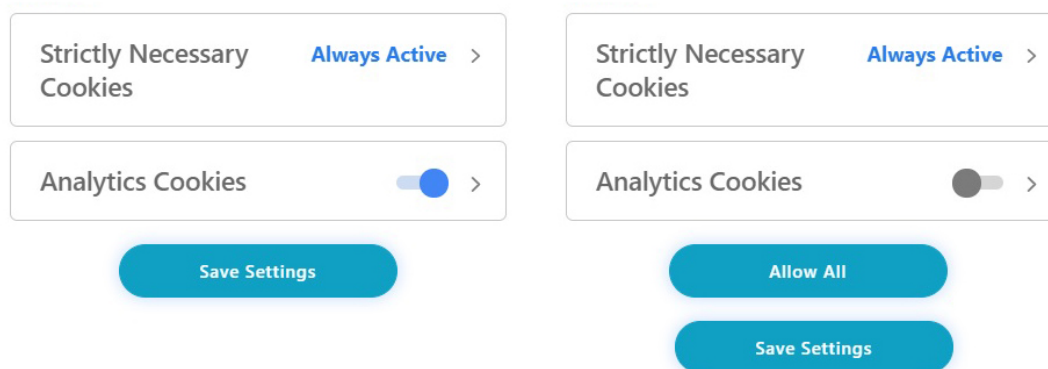


Figure 3.19: On the left is the cookie notice with “Analytics Cookies” selected, on the right the option is disabled, leading to the appearance of a new button “Allow All” in place of “Save Settings”.

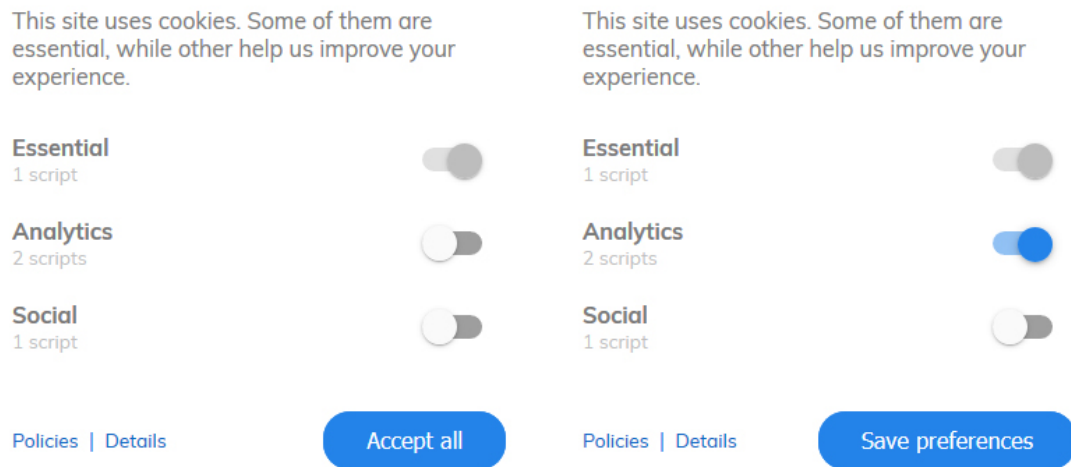


Figure 3.20: This privacy notification can only be closed by choosing “*Accept all*” or by selecting one additional option to “*Save preferences*”.

located. This additional button, which is unnecessary for saving the selection, entices fast clicking users to “agree” to more cookies even without noticing anything because the privacy notification closes immediately and the user may not have noticed that he/she clicked on a different button than intended.

Another example is a notice that cannot be closed, unless the user agrees in at least one category that cookies are set or scripts are executed, see Figure 3.20. On the left side of the Figure is the default option, with no pre-selected options but only allowing the user to “*Accept all*”. If a toggle is used to opt-in the button changes to “*Save preferences*”, but the privacy notification does not allow the user to close it and only set essential cookies. The cookie notice can be reopened after agreeing, and allows the deselection of the two options via “*Save preferences*”. However, this will cause the privacy notification to reopen and return to the initial situation, which results in the blocking of the content of the site and this may lead to the user consenting to the additional usage to close the notice to stop it blocking the content.

### 3.4 Dark Patterns and Nudging

Dark patterns describe a design method to use knowledge about human behaviour to deceive users to make a choice that is not in their best interest [48]. In the field of *Human-Computer Interaction* (HCI) this design approach is seen critical and raises ethical design questions, as the interest of the publisher takes priority over that of the user. While social norms determine what “good” is, the users should always be able to decide for themselves which option to take and not be tricked into making a decision.

In the following an overview of different types of dark pattern from [49, 50] are given.

### Trick Questions

In a form, a question is asked that tricks the user into giving an answer that was not intended. The question seems to ask one thing when viewed quickly, but on closer inspection it appears to ask another thing.

### Sneak into Basket

The user attempts to purchase one thing, but during the ordering procedure an additional item is sneaked into the shopping basket, to get the user to buy it. The item is often sneaked into on a prior site through an opt-out radio button or a checkbox.

### Roach Motel

The user finds it very easy to get into a situation or to create an account, but extremely difficult to get out. The opt-out is made very complicated to prevent the user from quitting.

### Privacy Zuckering

Named after *Facebook* CEO Mark Zuckerberg, the user is tricked into publicly sharing more personal information than intended. For example tricking the user to add a phone number for security reasons, but then made it available for all to see.

### Price Comparison Prevention

The user will find it difficult to compare the price of an item with another item, and thus make an uninformed choice.

### Misdirection

The user's attention is focussed on one thing in order to distract his or her attention from another aspect.

### Hidden Costs

During the last step of the checkout process, the user finds out that there will be additional costs, for example delivery charges or taxes.

### Bait and Switch

The user intends to do one thing, but is tricked into another undesirable option. For example the user wants to decline an upgrade, but is confronted with the options “*upgrade now*” and “*upgrade tonight*”, thus accidentally upgrading in an attempt to opt-out.

### Confirmshaming

This pattern guilt the user to opt-into something and words the decline in a way to shame the user to make the favourable choice.

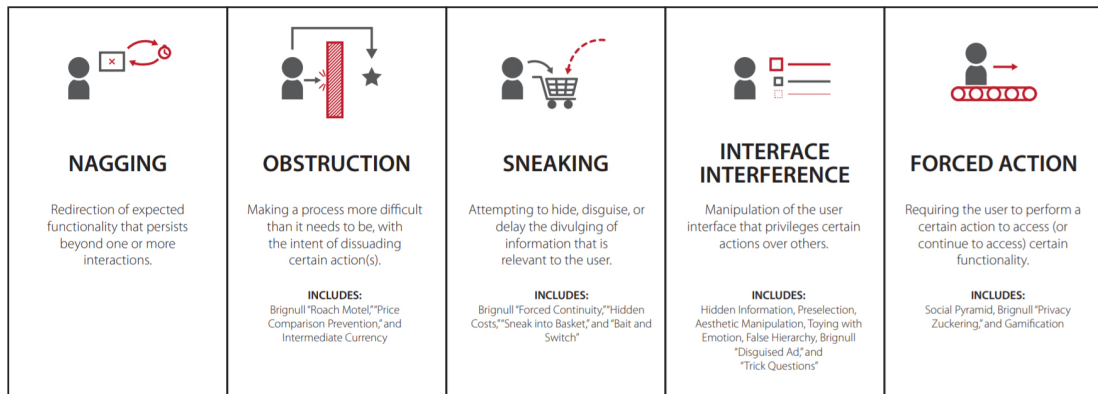


Figure 3.21: Dark pattern strategies, from [48].

### Disguised Ads

The user is tricked to click on advertisements, that appears to be part of the content or navigation.

### Forced Continuity

On a subscription-based service the user starts a free trial and has to enter credit card information. When the free trials ends and the user forgets to cancel, he/she is charged without warning or it is made extremely difficult to cancel the subscription before expiration.

### Friend Spam

The user is asked for an email or social media permissions under pretence of a desirable outcome, for example to find more friends, but instead all contacts are spammed with messages claiming to be from the user.

These dark pattern types can lead to strategies of nagging, obstruction, sneaking, interface interference and forced action [48], an overview can be seen in Figure 3.21. In section 3.1.1 different advertisements about the Brexit referendum are shown, that can be disguised ads, because the user may expect the ad to be an ordinary social media post or a survey through the options provided, but can be redirected elsewhere by clicking on the options in the advertisement. In the previous section 3.3, about privacy notifications, four of the five dark pattern strategies were observed.

### Nagging

In [48] nagging is described "*as a minor redirection of expected functionality that may persist over one or more interactions*". This strategy was not observed, an example of nagging is given in form of pop-up notifications or giving the user only the choices "*Not Now*" and "*Ok*", but not "*Disagree*" to stop the intrusion from



appearing again. Nagging could occur with privacy notification in the form of the cookie banner being displayed again and again, even if the user has already declined to make him or her finally agree so that the display stops intruding.

### Obstruction

The strategy obstruction describes a task that is unnecessary difficult to discourage the user [48]. An observed occurrence was, that the user visits the website to read the desired content and expects to close the privacy notification without additional consent. The simple interaction takes more steps than intended and interrupts the user in the intended task. Instead of clicking “Reject All”, for example the user must first open the settings and then open another eight sub-settings and withdraw the consent their individually. This will more likely result in the user agreeing to continue the intended task.

### Sneaking

“*Sneaking often occurs in order to make the user perform an action they may object to if they had knowledge of it*” [48]. This was observed in the form, that it was not always clear to the user how many companies would have access to the collected data or collect behavioural data themselves, and if the user notices he/she is faced with a lot of third parties that needed to be individually managed to deny consent.

### Interface Interference

This strategy was the most observed in [48] and describes the manipulation of the user interface, by deceiving the user to make certain actions, hiding information, for example pre-selections, or using design choices to misdirect the user. This strategy was observed in more privacy notifications. The user interface was manipulated in the form, that buttons changed their tasks or in the appearance of a new buttons to deceive the user. Information about the usage was hidden behind more steps and options were pre-selected, but only visible in the advanced settings. Furthermore, it was observed that the privacy unfriendly button was often highlighted and other options were hidden behind lighter colours or elements that appeared disabled.

### Forced Action

Forces Action is described “*as any situation in which users are required to perform a specific action to access (or continue to access) specific functionality*” [48]. Upon visiting a site the user is confronted with a privacy notification and has often only the options to accept the usage off cookies, as well as the tracking and selling of personal data or is excluded. This is often problematic in the form, that the user is required to use the online service, or on social media platforms, where the user is, after disagreeing also, excluded for part of today’s society.

A study about (un)informed consent [51] evaluated the interface of privacy notifications, highlighting the need for regulations regarding how this consent is obtained. In 2018 about

62% of popular EU websites displayed privacy notifications, but users were confronted with either too few or too many choices and developed a habit “to click any interaction element that causes the notice to go away instead of actively engaging with it and making an informed choice” [51]. From the investigated consent notices, 57,4% used dark patterns to nudge the user to make a choice not in their best interest. Examples are highlighting the privacy-unfriendly button, hiding the details or advanced settings and pre-selecting options, which were also found in the section 3.3. Publishers are more interested that user accept cookies and third party tracking and nudge users to make this choice.

## 3.5 Tools to Visualize and Block Tracking Methods

### 3.5.1 Firefox Lightbeam

*Lightbeam*<sup>1</sup> is a browser extension created with the goal to visualize and expose tracking from third party domains in real-time. Originally called *Collusion*, the Add-on allows insight in data collection of “the diverse range of third party companies that shape so much of our online experiences today from advertising to social sharing to personalization” [52].

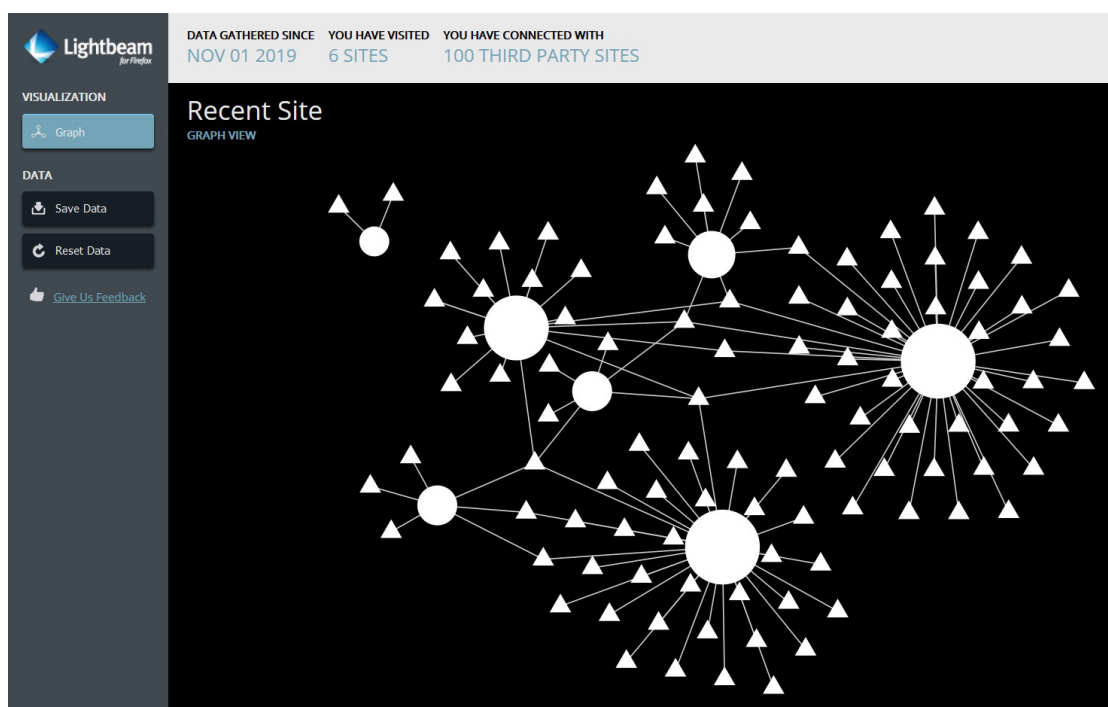


Figure 3.22: The Firefox extension *Lightbeam* after visiting 6 websites.

<sup>1</sup>Github site of *Lightbeam*: <https://github.com/mozilla/lightbeam-we>

After installing *Lightbeam*, the extension starts collecting third party cookie request for each site the user visits and displays the connections in a graph network, an example can be seen in Figure 3.22 after intentionally visiting six websites. Each circle represents a first party website directly visited and the size depends on the amount of connections to third parties. Triangles represent a third party domain the user indirectly connects with through a first party. The user can interact with the graph by clicking on a circle or triangle and the domain name of the first or third party is then displayed. Furthermore, it is possible to move single components by drag and drop to get a better overview. As the user continues browsing the graph grows and provides an overview of the online activities and tracking. Each time the user connects to a third party from a different website, the third party can access their own cookies containing already collected information and update them.

The tool was used to visualize the top 50 websites<sup>2</sup> from Austria. *Lightbeam* was started in a clean *Firefox* environment and each of the websites was opened without any interaction. The pages remained open for two minutes to establish all third party connections as well as cookie requests and the result can be seen in Figure 3.23. Only four websites show no cookie requests from third parties, whereas 10 websites had third parties cookie requests that only connect to their website. The others are all somehow connected and while *Lightbeam* allows drag and drop the graph is rather complex.

Overall *Lightbeam* counted 396 third party connections and in the *Firefox* cookie settings 125 domains with a total amount of 514 successfully set cookies could be observed. The highest number of cookies was from the first party news site *kurier.at* with 22 cookies.

Since no interaction has taken place on the websites, it was not confirmed that third party cookies can be set to enable tracking or analytic functions. However, the graph shows that many websites still try to set cookies. *Firefox* itself, depending on the privacy settings, can prevent the setting of third party cookies or the tracking of cross-site activity from social networks. This example shows that many third party cookies are already set by simply opening websites, which the user may not be aware of and *Lightbeam* shows the invisible privacy invasion from third parties.

### 3.5.2 Trackography

While reading online news a lot of information about interests and behaviour is revealed to third parties. Depending on the news sites and which articles were read personal data is collected where not all users are expecting them. The open source project *Trackography*<sup>3</sup> “aims to increase transparency about the online data industry by illustrating who tracks us online and where our data travels to when we access websites” [53], an overview can be seen in Figure 3.24.

*Trackography* focuses on global, national and local media websites as well as blogs and how information travels around the world. The user selects an available country and

<sup>2</sup>Top 50 websites from: [www.alexa.com/topsites/countries/AT](http://www.alexa.com/topsites/countries/AT)

<sup>3</sup>Explore *Trackography* at: <https://trackography.org/>

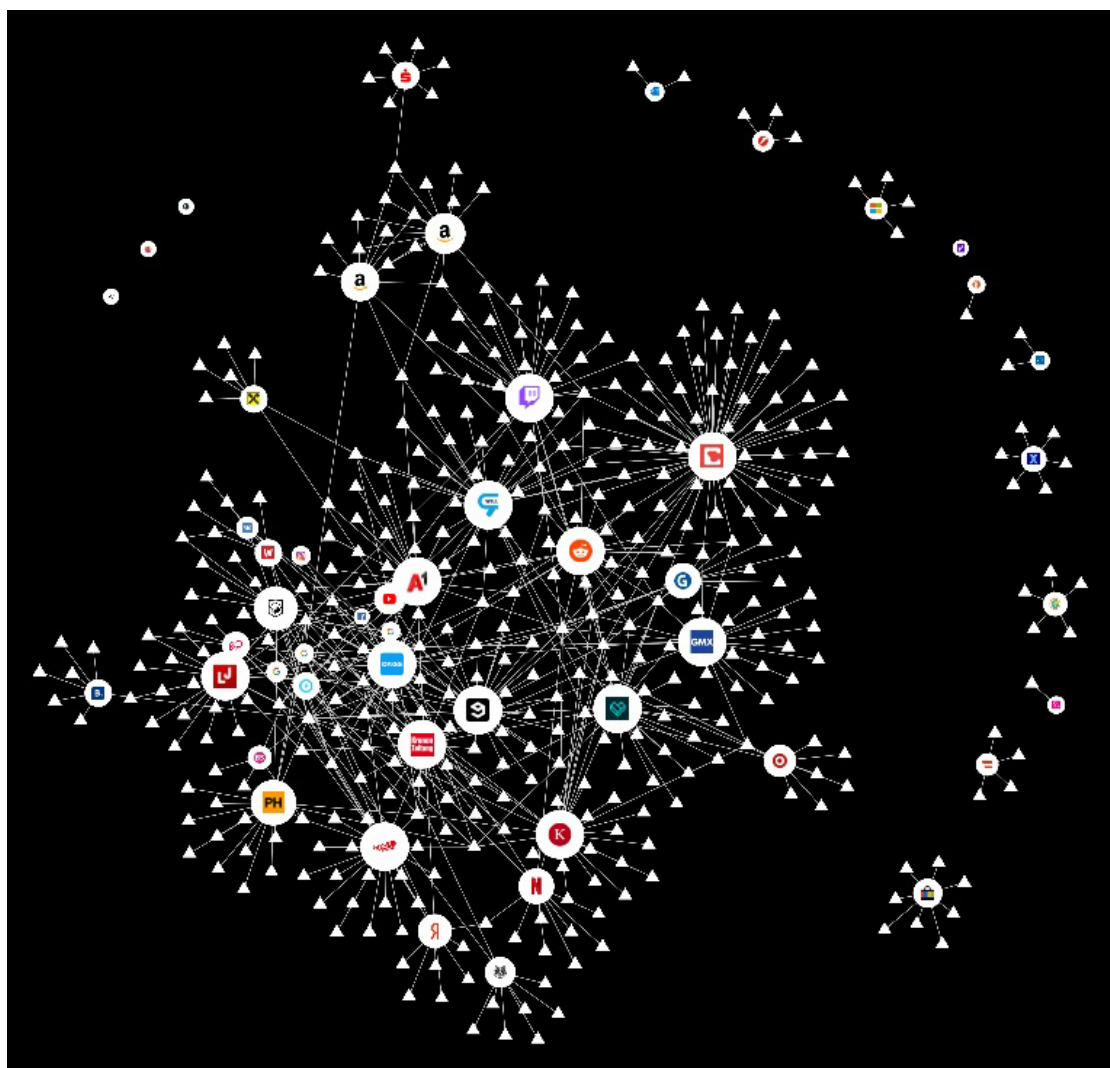


Figure 3.23: Austria's top 50 websites: third party cookies and their connections with no user interaction visualized with *Lightbeam*.

one or more news sites he or she is reading. The map shows an animation about where data is requested from and in which countries the servers are located. An example is seen in Figure 3.25 after selecting the local media *meinbezirk.at*. Connections to news sites are shown in blue and unintended third party connections in red. As origin country Austria is coloured in green, Germany is shown in blue, as the hosting servers for the website are located there. Countries in violet contain network infrastructure where a request is redirected and the countries containing a tracking service are shown in red. The coloured countries can be selected and additional information about the companies, like how long they collect the information and which laws they are subject to, are shown.

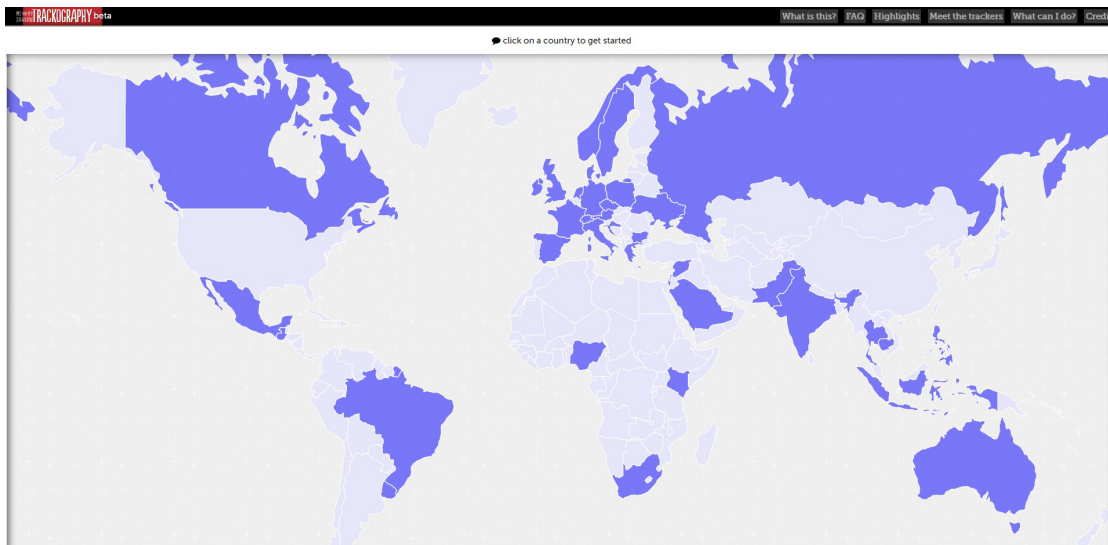


Figure 3.24: Overview of *Trackography* showing all available countries.

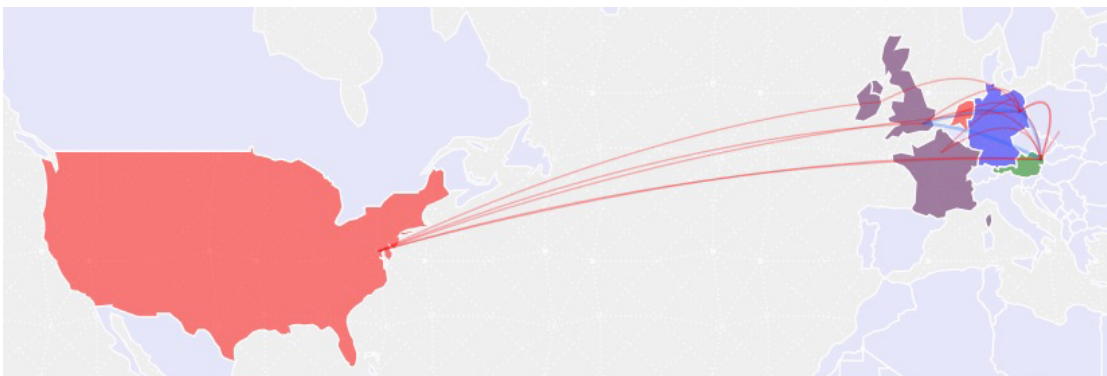


Figure 3.25: *Trackography* illustrates where data travels after visiting *meinbezirk.at*.

Overall *Trackography* shows 25 unintended connections to third parties after visiting *meinbezirk.at*. The project is still in a beta phase, because the data has to be collected in the origin country and is not displayed in real time.

### 3.5.3 Panopticlick

This project focuses on awareness about the tracking method fingerprinting and testing the browser of uniqueness. The tool uses accessible information about the browser and computer configurations and creates a fingerprint<sup>4</sup>. The uniqueness is evaluated from other users that used *panopticlick* in the last 45 days and the user finds out if the browser fingerprint is unique or how many others have the same fingerprint. Upon testing the

<sup>4</sup>*Panopticlick* fingerprint test: <https://panopticlick.eff.org/>

browser the user receives a report as well about how well the browser protects against tracking and how many other users have the same fingerprint as well as a list about the browser information that is accessible, for example browser plugin details, time zone, screen size, colour depth, system fonts, canvas fingerprint, language, platform, touch support or if an ad blocker is used. *Panoptlick* describes the methodology used and shows additional information to make the browser fingerprint less unique and the usage of the *Tor* browser is suggested to stop this tracking technique. It can be noted, that the more the users protect themselves with different browser extensions to block third party tracking, the more unique the browser fingerprint becomes.

#### 3.5.4 Tor Browser

*Tor* is a web browser that blocks tracking, resists fingerprinting and defends against surveillance [54]. The browser protects the user from traffic analysis, so they can not be identified by their IP address. By default, *Tor*<sup>5</sup> does not store cookies or browsing history, which are deleted when the browser is closed.

*“Our mission: To advance human rights and freedoms by creating and deploying free and open source anonymity and privacy technologies, supporting their unrestricted availability and use, and furthering their scientific and popular understanding.”* [54]

The browser is easy to use even for non-experts and is based on *Firefox*. Once installed and opened, a website is redirected using the *Tor Network* via three nodes to protect the users privacy and prevent tracking. By default, the *DuckDuckGo.com* search engine is used, which protects the users search behaviour and is not used for personalized advertising. The search results are not personalized, so every user gets the same search result. In addition, search queries are not stored and therefore cannot be used to identify user behaviour.

#### 3.5.5 Privacy Badger

*Privacy Badger*<sup>6</sup> is a browser extension available for *Opera*, *Chrome* and *Firefox* that “stops advertisers and other third-party trackers from secretly tracking where you go and what pages you look at on the web” [55]. The extension does not focus on blocking based on a list but it learns which trackers the user encounters while browsing. The first time a tracker is found *Privacy Badger* sends a *Do Not Track* request and if the tracker is found on another website ignoring that the user does not want to be tracked the extension blocks the tracker automatically regardless of their intention. The extension focuses on

---

<sup>5</sup> *Tor* project: <https://www.torproject.org/>

<sup>6</sup> *Privacy Badger* for *Firefox*: <https://addons.mozilla.org/de/firefox/addon/privacy-badger17/>

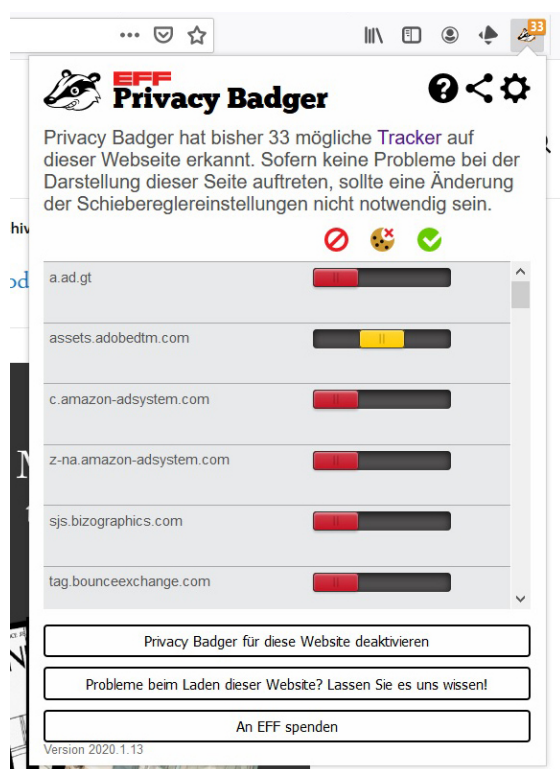


Figure 3.26: The browser extension *Privacy Badger* on the news site *newyorker.com*.

blocking different tracking methods like cookies, local storage, pixels or fingerprinting, but still allows third party content that is needed for the site to function.

In Figure 3.26 *Privacy Badger* can be seen in use on the news site *newyorker.com*, with 33 potential trackers. On the icon the number of potential trackers is shown and on click the extension expands for more details and to control individual settings. Is the slider red that means the tracker has already been observed and asked not to track the user, but the request was ignored and the tracker is now completely blocked. Trackers with the slider in the middle are yellow and wants to track the user but parts of the third party domain are needed for the functionality of the website. Green trackers have not yet been observed or have not yet attempted to track the user across multiple pages.

### 3.5.6 Ghostery

*Ghostery*<sup>7</sup> is a browser extension that, in addition to tracking, blocks advertising and thus protects the privacy of the user while browsing the internet. By blocking these contents,

<sup>7</sup> *Ghostery* for *Firefox*: <https://addons.mozilla.org/de/firefox/addon/ghostery/>

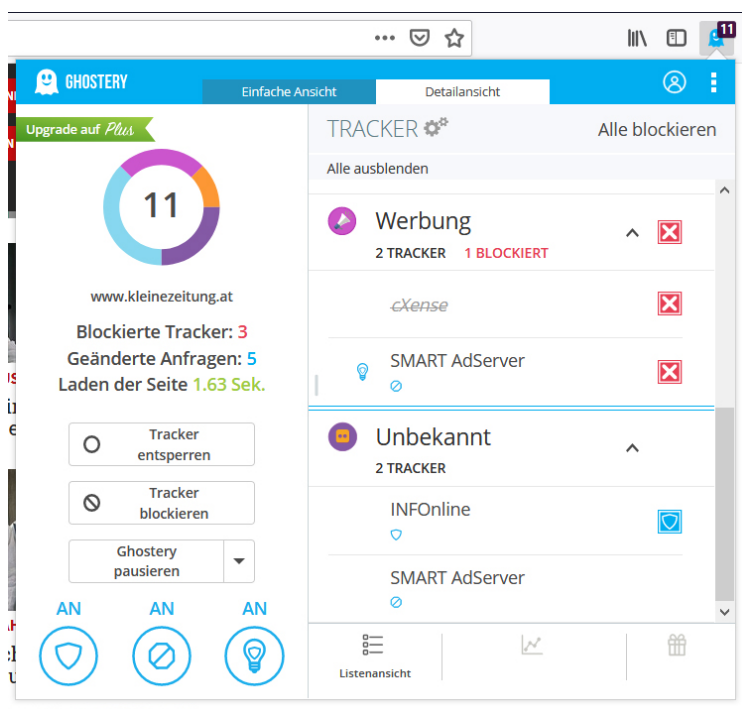


Figure 3.27: The browser extension *Ghostery* on the news site *kleinezeitung.at*.

websites load faster and are usually more user-friendly. In Figure 3.27 the extension is in use on the news site *kleinezeitung.at*, showing 11 trackers.

The extension divides the trackers into the following categories: essential, website analytic, advertising and unknown, which can be individually managed and blocked. The detected trackers can be managed and trusted or blocked based on the website, thus allowing the users to manage who collects data and to better protect their privacy.

By using ad blockers, the users can protect them self from the collection of personal data. However, from a publisher’s point of view is this not profitable, as it deprives them of a potential source of income and it can happen that the user is excluded from the use of a website if an ad blocker is used. But whenever an advertisement is loaded, not only does the user see it, but the advertiser connects to the user, getting the possibility to set cookies or use other tracking techniques to collect personal information.

### 3.6 Discussion

Every year *chiefmartec* publishes a landscape trying to capture all marketing companies and divides them into categories, like “*Advertising & Promotion*”, “*Social & Relationships*” or “*Data*”. In 2011 about 150 companies were researched who were active in this business,



however eight years later the supergraphic<sup>8</sup> includes 7.040 companies. Not all these companies are specialized in advertising technologies, but the graphic provides a good overview of how many companies benefit through the collection of personal data.

“*Personal data has become a new source of economic value*” [56], and provides companies with information about people’s behaviour and interests, which can be used for personalized advertisements. The two main data collectors are *Google* and *Facebook*, who share the personal information for marketing purposes with third parties. For example, *Facebook’s* advertising revenue was \$69,66 billion [57] in 2019 and *Google* recorded \$113,26 billion [58] in the same year worldwide.

An attempt was made to give an insight into the topic and to consider different perspectives. To show why it is important to deal with the topic some effects of adTech in the real world were described in the form of influences in elections. How adTech companies are able to collect personal data was shown in the previous chapter and in the form of privacy notifications, where users are often not aware of how many companies are collecting data in the background after accepting. At last some tools were shown, with which the users can protect themselves and already realized projects, which try to show privacy invasions.

Information collected by *Facebook* can be used for targeted political advertisements, that can nudge people to make a certain choice and influence democratic decisions. Problems are the source of the information and what goals are pursued by whom. Fake news addresses sensitive issues and are often based on a twisted truth, to divide the population and steer in a certain direction. Since personalised targeted advertising is only seen by the addressed audience, the potential influence remains longer unrecognised and prevents public discussions. Investigations regarding of potential influences are initiated after the election has already been carried out and it may take years before the influences are proven. The companies involved do not want to admit wrongdoing and only change policies under great pressure from governments and regulations.

The GDPR is a step in the right direction to protect the users privacy online, but it was shown, that adTech companies are always looking for new ways to follow the users unnoticed, with different techniques. The examples show that the implementation of the GDPR is only causally done on some websites, and the user only has the possibility to agree, is faced with obstacles to refuse consent or is nudged or tricked to agree, mostly without knowing it. Standards for privacy notifications should be introduced, for example, making it possible with a single click to “Reject All” tracking and collection of personal data, or possible browser solutions where the user makes a decision once and this decision is applied on each page visited, making it easier for the users to maintain their privacy online. Personal data is also provided directly by the user when using a first party website. By using their services, the user agrees to their privacy policies and the collection of data.

<sup>8</sup>The Supergraphic can be found on: <https://chiefmartec.com/2019/04/marketing-technology-landscape-supergraphic-2019/>

However, this is done by the average user with lack of information and without agreeing the users are excluded from part of today's society [56].

The user may be only seen as a tool for the adTech companies to increase their profits by tracking the users online behaviour. The publishers justify the use of tracked and personalized advertising on their websites to provide free content to the users. In a study [59] it was researched how users perceived personalized advertising relevance in relation to privacy concerns on social media platforms. It concludes that people with privacy concerns are more likely to avoid ads and *“worry that advertisers collect their personal information for marketing purposes”* [59]. If consumers perceived they are shown personal relevant ads they are more likely to show interest and pay attention to them, however if the advertisement is too personal people view them critically. This also shows the need to raise more awareness about the topic of advertising technologies and online privacy. That advertising does not always have to be personalized can be seen on the example of the *“The New York Times”*, that stopped behaviour targeted ads on their site to comply to the GDPR [60]. The website only displayed contextual and geographical targeted advertisements, that were direct-sold and not auctioned and did not lost any advertising revenue, showing there is an alternative way.

If a person looked at a product in a shop and decides against buying it and then someone would follow the person in other shops and even home, to ask if he or she wants to buy the product after all, one would consider this behaviour stalking. As not everyone wants to trade their privacy for the “most relevant” advertisement, this topic needs more awareness to stop the adTech industry from advancing.

# Analysis of News Sites

The previous chapters showed that the adTech ecosystem is a complex system about how to best track the users and auction for the right to show advertisements that are “relevant” to increase profit. But what one considers relevant can for another be misleading and influence choices by showing target advertisements. To protect the personal information and the privacy of EU citizens the GDPR came into force in 2018. Amongst other things publishers are now required to obtain consent on collecting personal information and enlighten the user how the data is collected and which third parties get access. Examples from exploring adTech show that not all publisher abide to the regulations and use tracking technology before consent is obtained or try to trick the user into accepting.

During the exploration of adTech newspapers have repeatedly attracted attention with their high number of cookies and trackers. The goal of this work is to find a way to enlighten on the topic and encourage awareness of the problems of targeted methods thus local online news sites were selected and analysed to explore the situation in Austria. This chapter focuses on the collection and classification of the data, as well as interesting findings on the news sites about privacy and advertisement. Whereas the next chapter explores how to visualize the collected information to encourage awareness of the problem of data collection and how important it is to protect one’s privacy.

## 4.1 Collecting the Data

There is a wide range of online news sources besides news pages, like social media sites, blogs or websites offering different services in addition to news coverage. In terms of accessibility and to reduce influence from other content the focus of the selection was only on news websites. The final choice was based on a current statistic of the 10 most popular news websites in Austria from September 2019 [61], see Figure 4.1. The most popular site is the website of the *Kronen Zeitung*, with a reach of 37% which is equivalent to 2,4 mio. unique users.

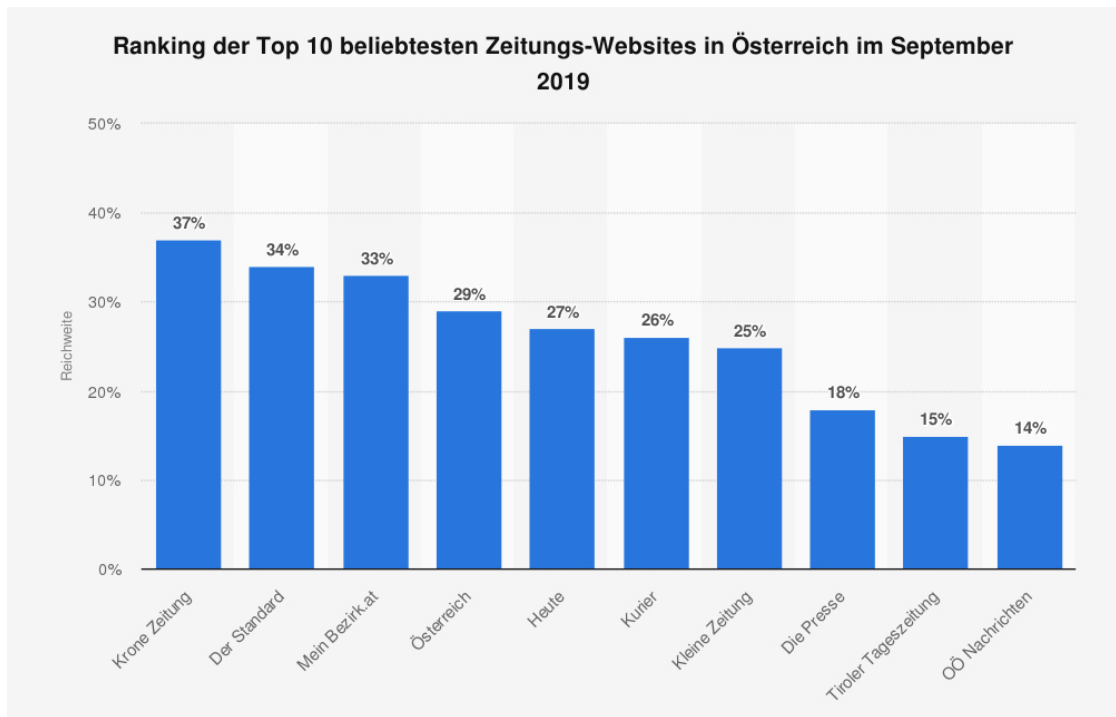


Figure 4.1: Top 10 of the most popular online news sites in Austria from September 2019, from [61].

The selected news sites with their corresponding domain name can be seen in Table 4.1. In the following the sites are referred to by their domain names and ordered alphabetically.

The data was collected with *webXray*<sup>1</sup>, “a tool for analyzing third-party content on webpages and identifying the companies which collect user data” [62]. The tool was created by Tim Libert and used in his work: “Good News for People Who Love Bad News: Centralization, Privacy, and Transparency on US News Sites” [63], to compare 4.000 US news sites with 4.000 non-news sites and found, that news sites relay more on third party content and risk the privacy of users and confirms the previous findings.

To start the analysis, *webXray* is given a list of websites, in this case the 10 domains of the news pages. For each site a new Chrome browser window is opened in a testing environment with a new temporary profile. This avoids any influence of the existing browser behaviour and already set cookies. While the page is loading all cookies and elements are logged and saved in an SQLite database. To collect all the data a loading time of 45 seconds is recommended, then the window closes and the next domain is opened in a new anonymous environment.

To show how much background activity is taking place that is not visible to the users three cases were tested. The first one is to find out if the publishers are following the

<sup>1</sup>Github site of *webXray*: <https://github.com/timlib/webXray>

News Site	Domain
Kronen Zeitung	krone.at
Der Standard	derstandard.at
Mein Bezirk.at	meinbezirk.at
Österreich	oe24.at
Heute	heute.at
Kurier	kurier.at
Kleine Zeitung	kleinezeitung.at
Die Presse	diepresse.com
Tiroler Tageszeitung	tt.com
OÖ Nachrichten	nachrichten.at

Table 4.1: Selected news sites with their domain names.

GDPR and obtain consent before tracking the user's and allowing third parties to collect information and set cookies. The second one is to simulate an average user, that just wants the privacy notification to go away, as found in section 3.4, and accepts the default settings. The last case is an extension of the second with the additional interaction of reading three articles to find out if tracking is increasing.

The following three cases were performed for each news site:

1. visit of the news site with no interaction
2. visit of the site and only accepting the default settings of the privacy notification
3. visit of the site and simulating a user: accepting the privacy notification, choosing a state if asked, visiting three random articles and scrolling to the bottom of the page to load all elements

For the first two cases the recommended time of 45s was used to capture the loading of all elements. The time was increased to 180s for the last case as three additional pages were loaded. The data was collected in the last week of November 2019. All three cases were carried out on the same day and verified a few days later with the same conditions. Table 4.2 shows an overview of the cases and the number of connections made by the news site to third parties as well as the unique connections shared by both datasets (a and b). For this comparison, only connections that loaded elements were used, third parties that only set cookies were not considered here.

	Case 1			Case 2			Case 3		
	a	b	unique	a	b	unique	a	b	unique
derstandard.at	1	1	1	11	12	12	36	47	49
diepresse.com	21	22	22	37	34	38	74	74	80
heute.at	18	18	18	20	18	20	57	61	74
kleinezeitung.at	20	24	24	34	31	36	73	69	75
krone.at	13	13	13	17	17	17	24	28	28
kurier.at	13	12	13	30	30	32	50	52	60
meinbezirk.at	14	14	14	28	20	28	45	46	58
nachrichten.at	14	14	14	24	24	26	31	37	39
oe24.at	51	45	52	49	56	63	70	75	86
tt.com	12	13	13	13	13	13	13	13	13

Table 4.2: Comparison of third party connections for each case and domain.

There are only small deviations in the number of connections in case 1 (no interaction) with the exception of *oe24.at* that connects to a high number of different third parties to load content. One additional domain was for example a *JavaScript* error troubleshooting domain or a domain from a company specialized in advertisement solutions.

After consent is given in case 2 and 3, the number of connections increases, except for *tt.com* which seems to connect to a constant number of third parties. For the other domains the number of connections are slightly different which can be due to the different content and advertisement displayed. This difference is greatest in case 3, where three articles were loaded randomly, which are different for both datasets. For example, if the article included social media content or other linked content then a connection to these elements was also established. Overall, all domains in both datasets and cases have approximately the same number of connections and there are no major differences that indicate errors in data collection.

Both datasets show only a snapshot of the content loaded and articles visited on one day. For the visualization only one dataset is needed, because it should be shown which connections to third parties are established and which content is loaded from them on visiting a news site once. Therefore, the first dataset is used for further analysis and the second dataset is only used for verification.

In Table 4.3 an overview of the third party connections, as well as how many first and third party elements were loaded, is given per case. In Table 4.4 connections to third parties who only set cookies are also included, which increases the number of connections.

		Case 1	Case 2	Case 3
connections	total	336	440	749
	unique	169	186	243
first party elements	total	665	1.542	2.156
	unique	664	1.531	2.068
third party elements	total	1.258	1.979	6.849
	unique	739	998	2.258

Table 4.3: Overview of total and unique connections as well as first and third party elements per case.

	Case 1	Case 2	Case 3
derstandard.at	1	13	52
diepresse.com	81	98	136
heute.at	18	20	81
kleinezeitung.at	105	107	131
krone.at	13	17	66
kurier.at	14	47	70
meinbezirk.at	14	28	54
nachrichten.at	14	24	51
oe24.at	64	73	95
tt.com	12	13	13

Table 4.4: The total amount of connections to third parties per news site and case.

## 4.2 Classification

The classification of the data was divided into two parts. First to identify the elements loaded and classifying them in different categories. The second part contains identifying the domain owners and the country in which they operate.

### 4.2.1 Elements

All content that was requested and cookies are regarded as elements. For each content-based element information such as the URL or original domain is stored. For the classification the elements were grouped if they share unique identifiers.

Grouped content-based elements have the same URL but differ in the query attached at the end. These elements may return different contents, but are counted as calls to the same element. For example, the following request of a pixel:

```
exampleAds.com/pixel?page=news1
exampleAds.com/pixel?page=news2&article=A
exampleAds.com/pixel?page=news2&article=B
```

These three calls count as one grouped element (`exampleAds.com/pixel`), because they share the same URL. With the attached query, in this example `exampleAds.com` knows that `news1` and `news2` have been visited as well as `articleA` and `articleB` have been called on `news2`. This information can be used, for example, to track the user's interests or to display specific ads based on information already collected.

Cookies count as grouped if the set name and the original domain match. For example, `exampleAds.com` tries to set or re-access the same cookie on three news pages.

News Site	Cookie Name	Owner Domain
news1.at	<code>_c_user</code>	<code>exampleAds.com</code>
news2.com	<code>_c_user</code>	<code>exampleAds.com</code>
news3.at	<code>_c_user</code>	<code>exampleAds.com</code>

In this case if the user visits `news2.com` the third party `exampleAds.com` can access the already set cookie `_c_user` again. This cookie can then be used to track the user and collect information to provide targeted advertisements.

In Table 4.5 the total amount of elements and the number of grouped elements per case is shown. Based on the extensions of elements *webXray* classifies elements already in *data\_structured*, *font*, *image*, *javascript*, *page\_static* and *style\_sheet*. With this option, a large part of the data could already be assigned. With the method of grouping elements,



the number of elements that needed to be classified by hand is reduced. All three cases loaded a total of 4.597 grouped elements and of these, 25.60% could not be clearly assigned to one of the above-mentioned types.

	Case 1	Case 2	Case 3	total
total Elements	1.923	3.521	9.005	14.449
grouped Elements	1.403	2.529	4.326	4.597

Table 4.5: Overview of total and grouped elements per case.

In order to assign the remaining elements, the following types were selected based on the existing classification and slightly altered to fit a broader group:

- **Cookie:** HTTP cookies
- **Media:** all forms of pictures, video and audio elements as well as sponsored content
- **Pixel:** elements the size of 1x1 pixel
- **Style:** CSS-files and fonts
- **Script:** JavaScript and PHP files
- **Structured Data:** hierarchical structures with attribute-value pairs, a whole list as well as one returned element
- **Static Page:** HTML files
- **Unknown:** elements that could not be classified

First elements with a file extension that were not already assigned to a type were classified. For example, all *woff2*-files, used for storing fonts, were automatically classified as style or *mp4*-files as media. Elements without file extension were mostly URLs with queries. In this case, the corresponding URL was accessed and the element was classified according to the content displayed. For grouped elements with different queries, three were randomly selected and if all loaded the same type all the elements in the group were classified this way. In some cases, different contents were displayed, for example two pixels and one image, then all grouped elements were classified as unknown, to avoid confusion after grouping the elements in the visualization. In most cases, this allowed a successful classification. For some elements, however, this was not possible because the elements were no longer found or the access authority for this element was not given and thus the access denied. These elements were also classified as unknown.

A special classification was that of tracking pixels. These are mostly elements with queries or gifs and therefore classified as image. As a common tracking method to collect

online behaviour or to match targeted advertisements it was interesting to know how many are in use. Most of them are not visible at first glance when opened, because they are only one pixel in size (see Figure 4.2), but therefore easier to classify. Usually the name of the element gives them away, as they are often called: *1*, *1x1*, *pixel*, *blank*, *collect*, *info*, *match* or *track*.

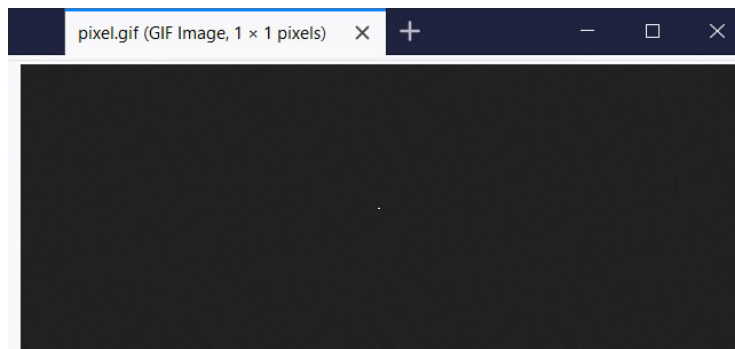


Figure 4.2: Example of opening a tracking pixel.

With this method most elements could be classified, except for 299 grouped elements (6,50% of total grouped elements) which were called 1.799 times (12,45%). The results of the classification are shown in tabular form for each news page, separated into the respective cases and if the element originated from the news site (1<sup>st</sup> party) or loaded from a third party (3<sup>rd</sup>rd).

The tables are annexed, for *derstandard.at* (Table 1), *diepresse.com* (Table 2), *heute.at* (Table 3), *kleinezeitung.at* (Table 4), *krone.at* (Table 5), *kurier.at* (Table 6), *meinbezirk.at* (Table 7), *nachrichten.at* (Table 8), *oe24.at* (Table 9) and *tt.com* (Table 10).

#### 4.2.2 Domains

The next step was to identify the domain owners. *WebXray* includes a domain owner list, with the most common third party domains. Further information includes the owner company, their headquarters location and if present a parent company. This information is mapped on the collected data and an example of the domain *googleadservices.com* looks like this:

Domain	Owner	Country	Lineage
<i>googleadservices.com</i>	AdSense	US	AdSense>Google>Alphabet

All three cases combined connected to 255 unique third party domains, from which 125 (49,02%) were not included in the domain owner list. For these unknown domains a WHOIS <sup>2</sup> query was executed. From this query the registrant organisation was taken as owner and the registrant country as the owner country.

<sup>2</sup>WHOIS is a protocol to store domain information and a WHOIS query allows users to look up this information in a readable form: <https://www.whois.com/whois/>

 <b>Registrant Contact</b>	
Name:	WhoisGuard Protected
Organization:	WhoisGuard, Inc.
Street:	P.O. Box 0823-03411
City:	Panama
State:	Panama

Figure 4.3: A privacy protection company is listed instead of the owner information of *cleverpush.com*.

A special case were domains where the owner could not be determined. Some companies disclose the registrant contact information, either by opting for privacy, then the WHOIS database only returns *REDACTED FOR PRIVACY*, or by using a third party company in their place to conceal their identity. An example is seen in Figure 4.3, from the WHOIS query about *cleverpush.com* that is using a privacy protection company located in Panama.

In the first run instead of the missing domain owner the registrar was used. This led to confusion as domain name registrars, for example *GoDaddy* or *Network Solutions*, manage the registration of the domain, but are not responsible for the content. For the visualization it was considered more important to show, that the domain owners did not want others to know who they are.

This led to classifying domains that could not clearly be assigned to an organisation as *Unknown* instead of using the registrar as the owner. For domains marked as *Unknown*, the information about the owner country was also missing. In this case the country of the server location was used to give a general overview of the origin. Researched owner information can be distinguished from the owners identified by *webXray*, that the lineage information is missing in the data. They were not research for this amount of data and assuming that the companies and their lineage listed in the *webXray* database are the most occurring ones in the tracking business.

In case 1 (see Table 11), visiting the news site with no interaction, the top 5 domains are from *Google*, now knowing the user has visited 9 of the 10 observed news sites. In case 2 (see Table 12), accepting the privacy notifications, *Google* is present on all 10 news sites along with *INFOnline GmbH*, and the top 14 domains are present on more than half of the observed news sites. And finally, in case 3 (see Table 13), accepting with interaction, 8 domains are present in all news sites, of which 6 are owned by *Google*.

Because most of the top domains belong to *Google (Alphabet)* in Table 4.6 an additional

ranking of the top 10 company owners can be seen, combining all occurrences of their owned domains. Some third party connections are necessary for the website to function, e.g. style elements or scripts, but can thereby also already collect information. Researching these companies listed in the Table revealed they are all specialized either in advertising, marketing or analytic technologies and connect to the user before giving consent of potentially tracking or the permission to collect analytical data.

Case 1	Case 2	Case 3
9 Alphabet	10 Alphabet	10 Alphabet
8 INFOnline GmbH	10 INFOnline GmbH	10 INFOnline GmbH
6 Adition Technologies	7 Adition Technologies	9 Adition Technologies
5 AppNexus	6 AdForm	9 AppNexus
4 comScore	6 AppNexus	9 IPONWEB
3 Taboola	6 Yieldlab	9 Media Math
3 AdForm	5 comScore	9 The Trade Desk
3 Adobe Systems	5 IPONWEB	8 AdForm
3 Beeswax	5 Facebook	8 Adobe Systems
3 IPONWEB	5 Media Math	8 Facebook

Table 4.6: Top 10 third party owners and their number of occurrences on the 10 observed news pages per case.

### 4.3 Other Findings

Originally, a fourth set of data was planned, in which all tracking methods and personalized advertisement were to be rejected, except for the essential functionalities. However, as seen in Table 4.7, only four news sites offer this option. For a future analysis, it would be interesting to collect this data, as well as in comparison with the option to pay not to sell data, in order to determine how these options affect the privacy of the user.

An example of what a pay wall can look like is shown in Figure 4.4 of *derstandard.at*. This gives the user the choice of viewing the website for free and therefore accepting personalized advertising and tracking, or paying for subscription and getting access without advertising.

A privacy notification, which does not allow any adjustments directly in the notice, can be seen in Figure 4.5 of *oe24.at*. The notice nudges the user to click on “*ok, understood!*”. However, if the user reads the privacy declaration, the possibility to revoke the consent of the setting of cookies can be found under “*Cookies*”, which reads like it is only possible

	Offers Privacy Adjustments	Offers to Pay for no Ads
derstandard.at	no	yes
diepresse.com	yes	no
heute.at	yes	no
kleinezeitung.at	yes	no
krone.at	no	no
kurier.at	no	yes (but two advertisers)
meinbezirk.at	yes	no
nachrichten.at	no	no
oe24.at	no	no
tt.com	no	no

Table 4.7: Overview if the news sites offer privacy adjustments or payment for ad free content.

**Ich stimme der Verwendung von Cookies für die Zwecke der Webanalyse und digitaler Werbemaßnahmen zu. Auch wenn ich diese Website weiter nutze, gilt dies als Zustimmung.**

Meine Einwilligung kann ich [hier](#) widerrufen. Weitere Informationen finde ich in der [Datenschutzerklärung](#).

OK

#### Abo ohne Daten-Zustimmung

Mit einem derStandard.at PUR-Abo kann die gesamte Website ohne zustimmungspflichtige Cookies und ohne Werbung genutzt werden. [Details zum Abo.](#)

PUR-Abo

🔑 Anmelden / Community / Sales / Über uns / Impressum & Offenlegung / AGB

Figure 4.4: Pay wall of *derstandard.at* offering to accept cookies or pay to view the content ad free.

after consent. Including this option directly in the privacy notification would probably result in a higher opt-out rate, which would not be beneficial to the publisher.



Figure 4.5: The privacy notification of *oe24.at* nudges the user to accept third party cookies.

The sites *kleinezeitung.at* and *diepresse.com* appear to use the same privacy layout, as seen in Figure 4.6 and 4.7. In both, five different purposes can be selected and individually managed. Furthermore, it is possible to manage all third party companies that have access to the data individually. The site *kleinezeitung.at* has 466 listed third party companies and *diepresse.com* has 468 companies. Finally, after making the selection, the user is tempted to click the highlighted button “OK, accept all” or “Accept Cookies”, which would lead to accepting all usage and would make the individual de-selection invalid.

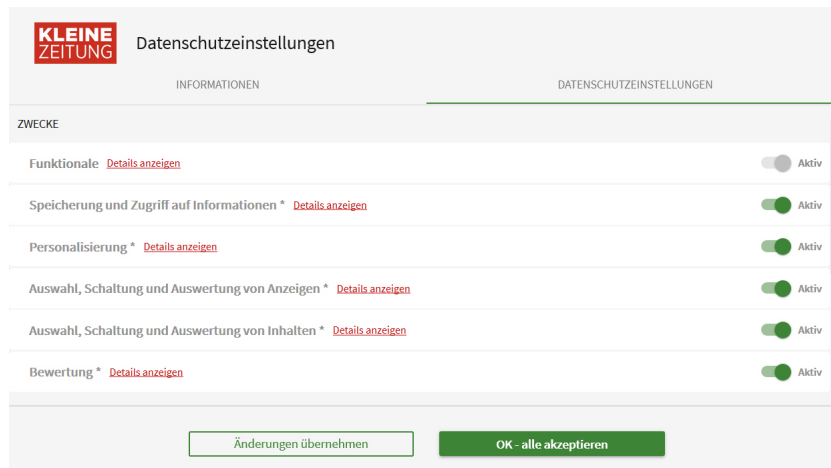


Figure 4.6: The privacy settings of the privacy notification of *kleinezeitung.at*.

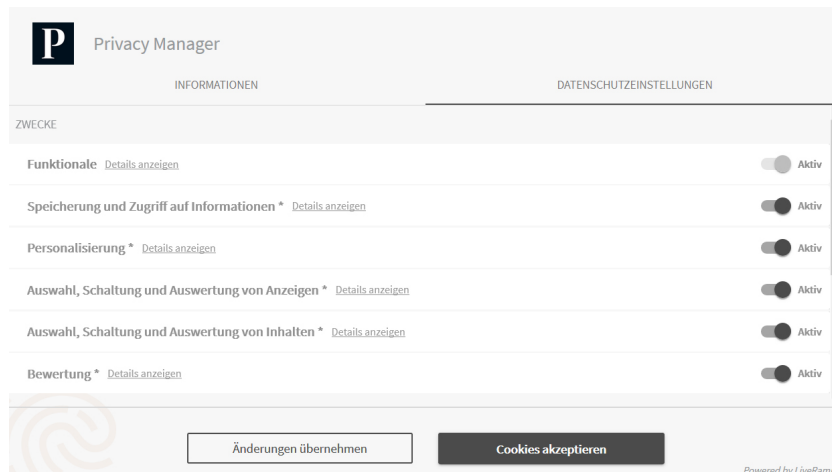


Figure 4.7: The privacy settings of the privacy notification of *diepresse.com*.

Advertisements are shown in different ways on the news pages, for example at *heute.at* one is surrounded by advertisements when reading an article, an example can be seen in Figure 4.8.

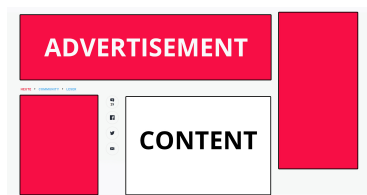


Figure 4.8: Advertisement in relation to content on *heute.at*

Another form of advertising is sponsored content and advertisements that look like news articles and can only be distinguished from other articles by the addition of “*paid advertisement*”, “*advertisement*” or “*sponsored*”. Most of the time these ads can be found under the section “*this might interest you*” and are sponsored by an adTech company. Examples of such advertisements can be seen in the Figures 4.9, 4.10 and 4.11.

Looking at this sponsored content from different news pages they have in common that they think I am an elderly retired person who needs a stair lift, a new shower, miraculous exercises and tips against joint pain. Since these screenshots were taken in the same test environment in which the data was collected, the only targeting could be location based. However, since “*Berndorf*” is not in the near vicinity, it could not be found out why targeted advertisements from three news sites show the same topic.

Some advertisements are structured in the form of clickbaits and contain headlines that entice the user’s to click on the ad because they want to know the rest of the information, for example “*number 1 trick to...*” or “*simple tips relieve years of pain ...*”.

#### 4. ANALYSIS OF NEWS SITES

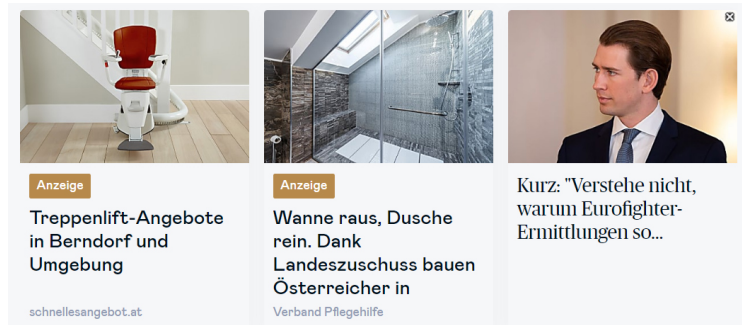


Figure 4.9: Sponsored content from *diepresse.com*, including two advertisements and one news article.

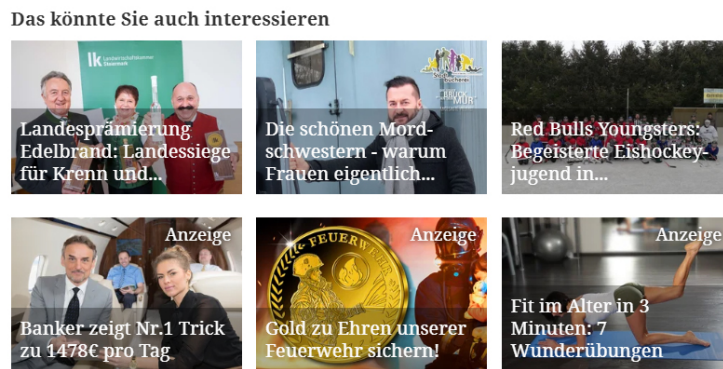


Figure 4.10: Sponsored content from *meinbezirk.at*, including three news articles (top row) and three advertisements (bottom row).



Figure 4.11: Sponsored content from *oe24.at*, including one advertisement and one news article.



# Interactive Design

This chapter focuses on finding out how to visualize the collected data from the previous chapter in a way that makes connections, that are established in the background, visible to the user. The user should be able to interact with the visualization in order to gain a better understanding of which companies load elements on which news pages, and thus potentially collect information. In order to see how the number of connections changes and which elements are additionally loaded after the privacy notification has been approved, it should be possible to choose between the collected cases.

First the different prototypes that have been implemented are discussed and why they proved no satisfying way to visualize the data. These prototypes lead to the idea of the final visualization which is discussed next as well as the different elements, how the user can interact with the visualization and finally further improvements are evaluated.

## 5.1 Design Prototypes

To show the third party connections for each news site and how they are connected different design ideas were implemented. The aim was for the user to interact with the data to gain a better understanding and insight about what elements are loaded and how many third party connections are taking place in the background. Existing visualizations, for example *Lightbeam* (see Figure 3.23) use a networking graph to show the connections, but it can be complicated quickly and thus losing the overview.

The visualizations were implemented using the *JavaScript* library *d3.js*, which stands for *Data-Driven Documents*. The library is used for making interactive visualizations in web browsers and allowing the manipulation of big data files. The script is embedded, and allows to draw vector based elements within an SVG, which can be scaled freely depending on the user's resolution. An important reason why this library was chosen was the fast loading of large amounts of data and the flexible display of these data. When a

new data set is loaded, D3 provides easy ways to update existing data. It is possible to map elements from different data sets to the same displayed element and thus perform a smooth transition of the data without redrawing all elements.

In the following the design process with implemented prototypes are presented leading to the final visualization.

### 5.1.1 Sankey

The first idea was to show how many connections each website has and the lineage information about the third party companies, a concept can be seen in Figure 5.1.

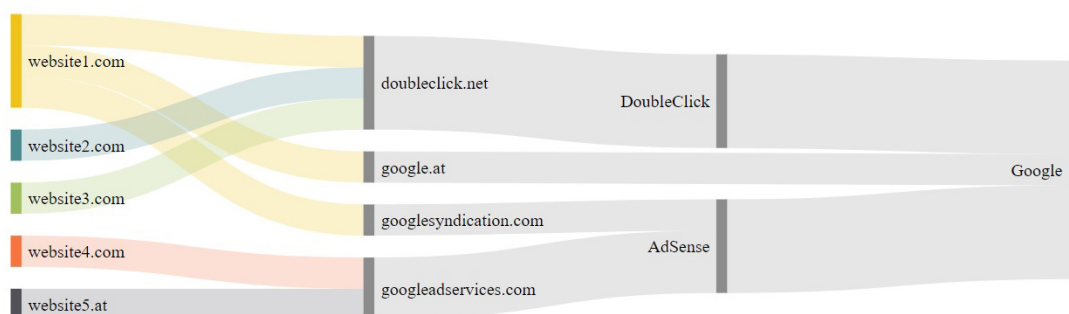


Figure 5.1: Concept of a Sankey visualization showing third party connections and third party lineage information.

The implementation of the concept with a small prototype dataset of five news pages can be seen in Figure 5.2. The news sites have a distinguishable colour and each step from the first party domain to the last owning company of a third party is represented by a node. A connection from the news page to a third party is realised with a link. The thickness of a link after a third party domain indicates how many news pages connected to this domain. To show where data travels the third parties are colour coded by country, for example the United States are blue. To make the visualization clearer, node descriptions with two or less outputs were removed, but can be viewed by mouse over. The description of the last known node is always shown and depending on the amount of connections the text size increases.

The user can interact with the visualization by hovering over a node or link for more information about number of connections, owner, lineage or country. To investigate a flow, it is possible to click on any node and all incoming and outgoing connections and nodes are highlighted, an example can be seen in Figure 5.3 after clicking on the adTech company *DoubleClick* and hovering over the link to the domain *doubleclick.net*.

This design was not further improved as it was getting rather complex with the whole dataset, especially the one from case 3 with 749 connections from news pages to 243 unique domains. The visualization became quite long and one had to scroll a lot to see

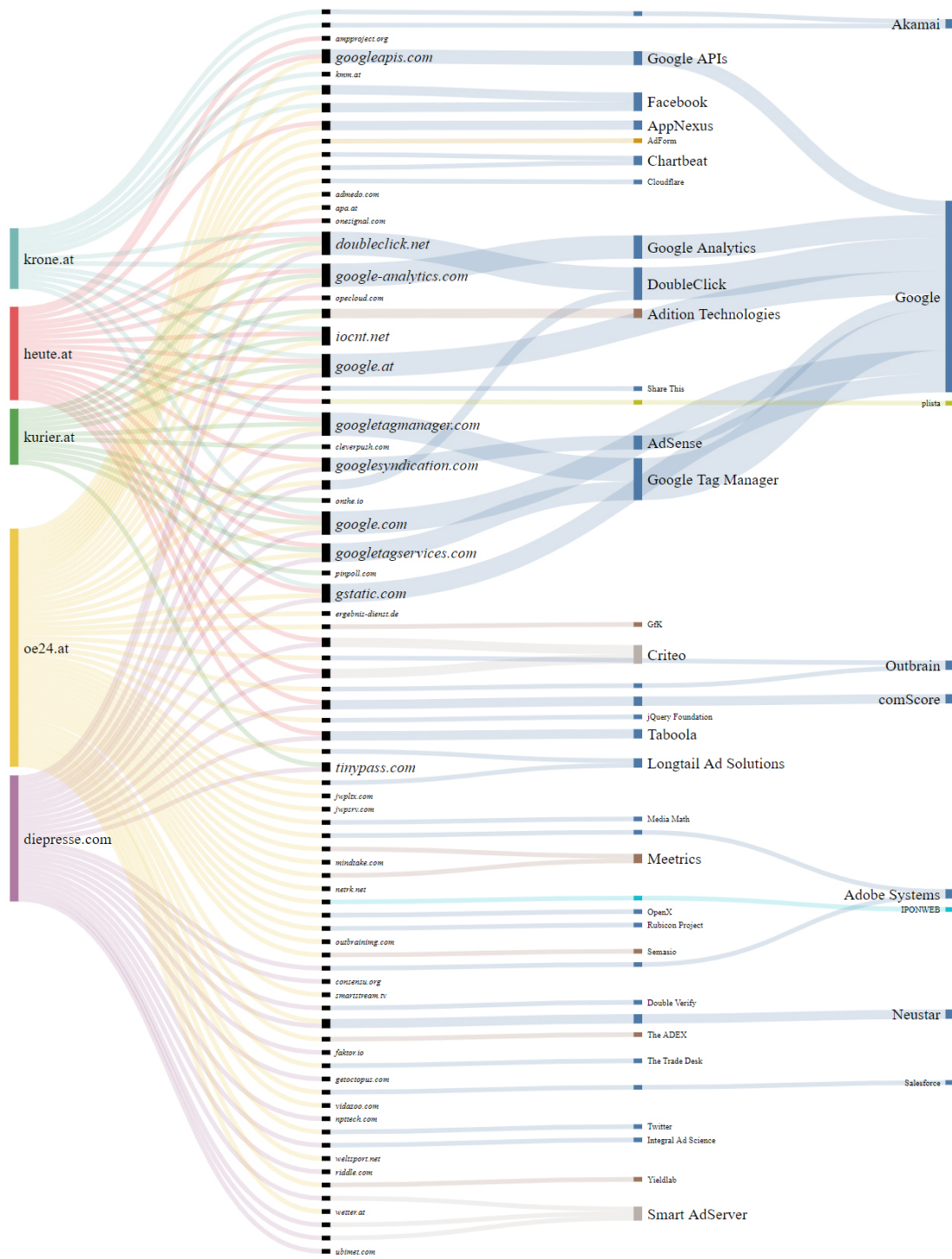


Figure 5.2: Prototype implementation of a Sankey diagram showing the flow of third party connections for each news page.



Figure 5.3: Section of the Sankey, showing the highlighting of the third party company *DoubleClick*.

all the information and quickly lost the overview. *Google* appeared to be the main player and connected to most subsidiary companies showing the most information about lineage.

### 5.1.2 Circle Packing

To concentrate the information a new design approach was realized that represents each newspaper by a circle. The size depends on the amount of connections and the circles are sorted according to size. Every connection from the news page is placed inside the circle as a dot, coloured by country. To make the best possible use of space and reduce empty areas, a phyllotaxis pattern<sup>1</sup> was implemented instead of the default placement, a comparison can be seen in Figure 5.4.



Figure 5.4: Comparison between the default pattern (left) and phyllotaxis pattern (right).

The user can choose between the three datasets (case 1 to 3) and the circles would resize based on the number of connections and change their position. On hovering the circles or dots additional information is displayed. The user can interact with the visualization by selecting a domain, then all domains belonging to the same company would be highlighted everywhere to show on which news sites the company can be found. An example of a

<sup>1</sup>A phyllotaxis pattern describes a spiral like pattern inspired by nature, for example the arrangement of the seeds of a sunflower.

prototype implementation of this concept can be seen in Figure 5.5, highlighting the domains connecting to *Google* on all news pages, on visiting with no interaction.

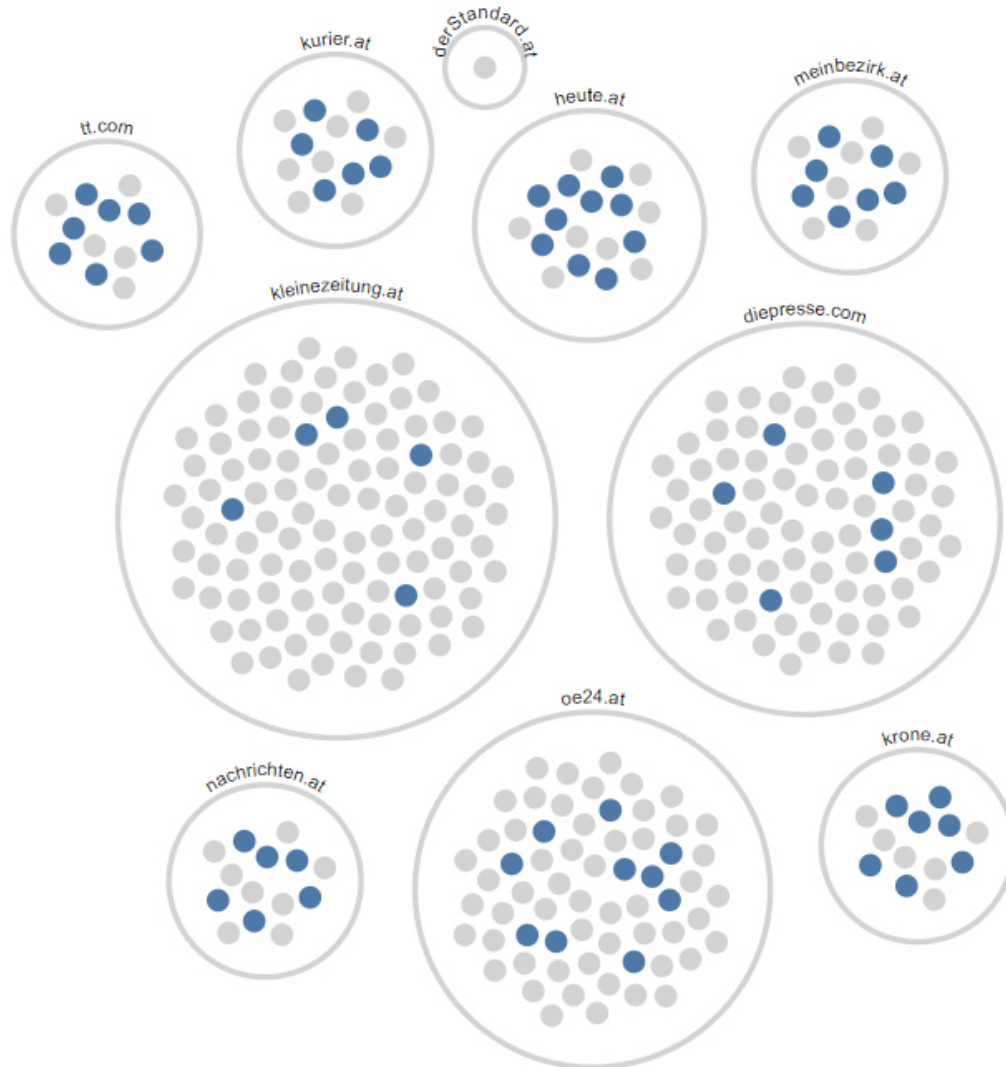


Figure 5.5: Circle packing prototype showing the number of third party connections for each news page.

Working on this concept sparked the idea to not only show the connections, but it would be interesting to know what elements each website loaded, leading to the classification of elements. Each element is now represented as a dot and each different type is represented by a symbol. By switching between the datasets, the elements rearrange themselves and one can see which elements remain the same and which new ones are added. Figure 5.6

shows all three datasets of *derstandard.at* side by side. First the elements are highlighted by type and first party elements are displayed a little brighter to distinguish them from third party elements. The bottom row shows the elements highlighted by country and sorted that only third party elements can be examined.

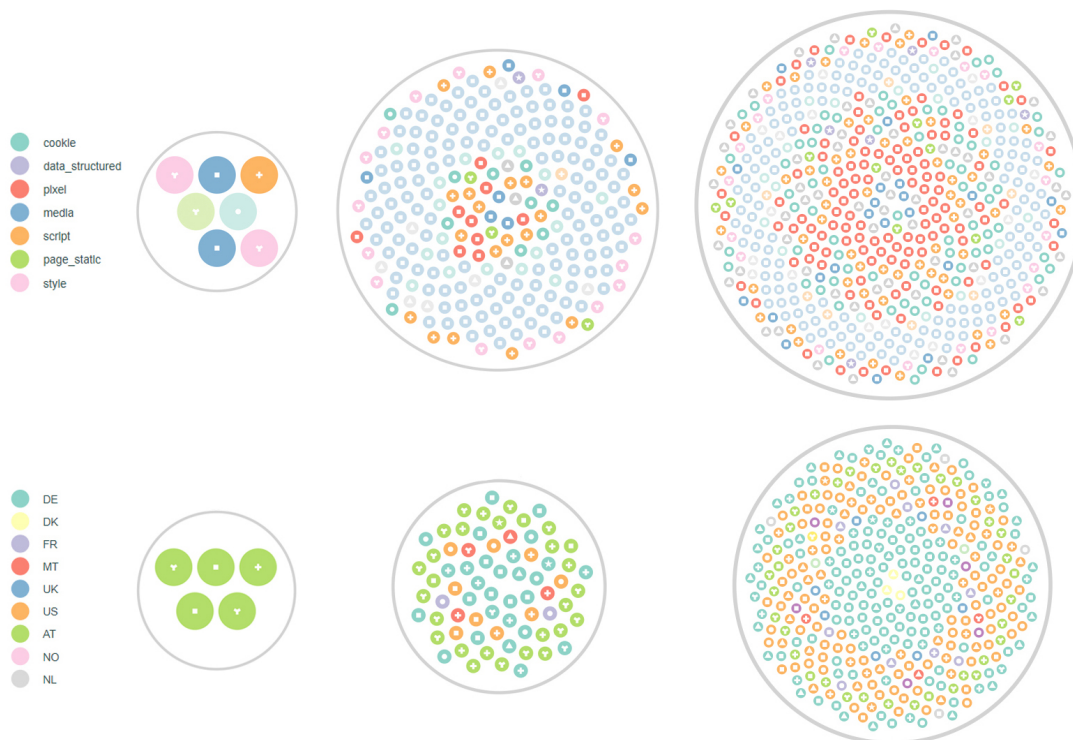


Figure 5.6: Prototype of showing all loaded elements inside a circle for each case, coloured by type (top) or country with only third party elements (bottom).

All in all, this prototypical visualization showed how many elements each page loaded and which new elements are added depending on the case. To simplify the interaction and to better analyse the data, the possibility to filter or select by type, country or company would have been added. But this implementation led to the problem that in case 3 with about 9.000 elements one would have lost the overview and the dots would have been too small on a standard browser resolution. However, if only one news page were to be displayed at a time, the elements could be better examined, but the information about the connections from third party companies to each news site would be lost.

### 5.1.3 Final Visualization

The main problem to find a suitable form of visualization was the amount of data. The circle prototype of Figure 5.5, was satisfying in the way that all third party connections could be displayed clearly and with the interaction one could find out with how many

domains and how often a company was represented on the news pages. Since it should not only be about showing how many third party connections are established, but also about showing possible tracking and collection of personal information, it was also important to know which third party elements are loaded. For example, a domain that only loads images can also collect data about the user, but a domain that already sets 30 cookies or calls up a pixel 18 times with different queries without user interaction when a page is accessed, raises more suspicion and poses a higher risk for the user's privacy.

This lead to brainstorming new design possibilities, which show the connections of news sites to third parties, but also which element types have been loaded, as well as on how many news pages a third party company is found on and whether the same elements are loaded there. In Figure 5.7 sketches of possible implementation variants can be seen.

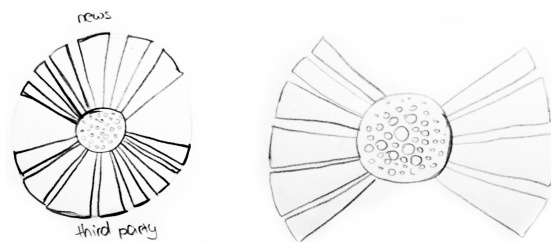


Figure 5.7: First sketches of the final visualization.

## 5.2 Implementation

The final visualization is implemented with the *JavaScript* library *d3.js* as well, as already developed elements and procedures, such as loading the elements, placement, update patterns and interactions have already been developed in the prototype variants. D3 offers a chord diagram (d3-chord) that could serve as a basic structure. This diagram arranges elements in a circle and shows dependencies as a flow between the elements inside the circle. After further search a graph was found, which is based on d3-chord, but has the functionality of the divided halves, as well as the possibility to manipulate single links, the d3-loom by Nadiyah Bremer<sup>2</sup>.

This plugin will be used to display the individual news pages and third parties, as well as the connections. The following information should be visualised:

1. the difference between the three data sets (cases)
2. how often an element has been loaded and its type
3. compare which elements two news pages have in common
4. how many third party connections a news site establishes and to which companies

<sup>2</sup>Github plugin site: <https://github.com/nbremer/d3-loom>

5. how many and which domains a company has
6. which element types are loaded the most, by news page or company

The first prototype of this idea can be seen in Figure 5.8 and showing a way to display the information more centred on a screen. On the left side the 10 news sites are displayed, where the thickness of the link leading to the circle indicates how many third party connections the user makes by accessing this page. In the middle, the circle represents the user and the individual dots inside, the elements the user has come into contact through the visit of the news pages. From the middle links lead to the right, which represents the third party domains. The thickness of these links indicates from how many news pages a connection to this domain has been established and therefore have a maximum thickness of 10. The links of the domains are then grouped by a bar if they have the same owner.

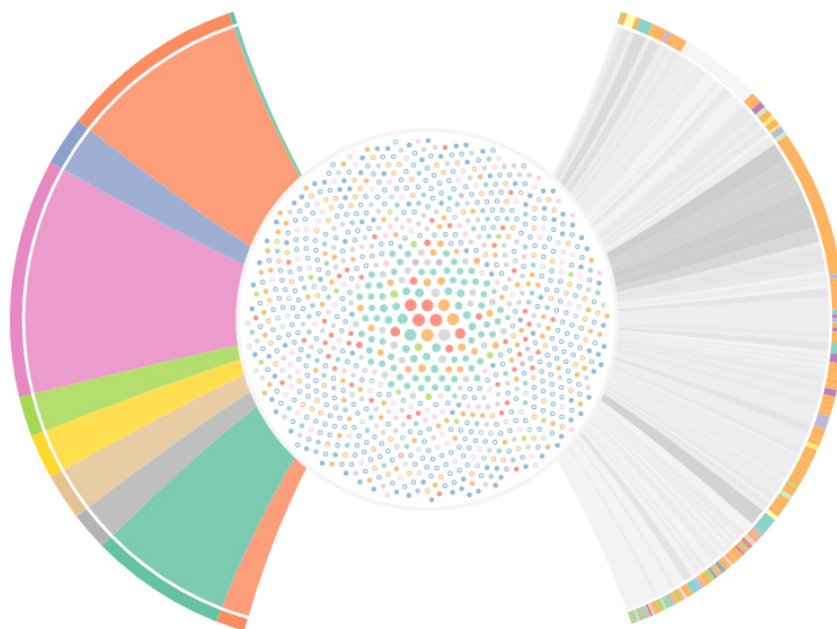


Figure 5.8: First working Prototype showing the connections of the news sites on the left, all loaded elements in the middle circle and the third parties on the right.

The classified elements of the three data sets are stored in JSON format and are loaded according to the case selected. Each call to an element that has been recorded has an entry with a unique ID which consists of the hash value of the URL for a content-based element or the hash value of the name and the associated domain for cookies. This ID has been added to prevent elements from the same domain that have the same name but a different path (URL) from being grouped incorrectly. An example of an entry looks like this:



```

{
  "page": "derstandard.at",
  "id": "6a7108d3f44a7f2c0fad96832c3f964b",
  "name": "privacywall-a749c49671.js",
  "type": "script",
  "domain": "staticfiles.at",
  "owner": "Standard Verlagsgesellschaft",
  "country": "AT",
  "lineage": "",
  "is3P": 1
}

```

Depending on the information required, the entries are filtered and grouped. If, for example, the number of connections per page is to be determined, the data is grouped according to page and domain and filtered to `is3P`, thus the number of different third party connections per news page can be obtained.

### 5.3 Colours

In the first prototype (see Figure 5.8) it was already experimented with colours and it was important to find the right colour balance, that the graph would look clean and not overloaded.

A colour palette from *D3* was used for the element types, *d3.schemeSet3*, which provides twelve colours for categories that are easy to distinguish. The eight element types and their colour are based on the scheme set and can be seen in Table 5.1.









			
cookie	pixel	media	script
			
style	static page	structured data	unknown

Table 5.1: Element types and their assigned colour.

Originally each news page had its own colour and the third parties were coloured by country. For the news sites the colours brought no additional information and for the third parties the information was overloaded, since the arc bars are already so small due

to the number of companies. Therefore, more intense colours than the elements were chosen for the arc bars (see Table 5.2), either red for news sites or blue for third party companies, to match the grey for the connections.

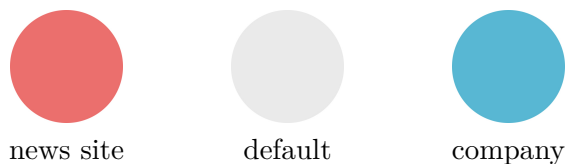


Table 5.2: Colour scheme for arc elements.

Different shades of grey were used to better distinguish the connections. The gradient can be seen in Figure 5.9. Depending on how many connections a domain has, the colour is interpolated, starting with one connection (left) up to a maximum of 10 connections (right).



Figure 5.9: Colour gradient for the third party connections.

## 5.4 Labels

One challenge was to put all names of the third party companies on the graph. In the largest data set of case 3, there are 243 connections to 149 different companies, which needs to be displayed.

In order to increase the readability different solutions were tried. First, connections that only led to one news page were combined, but this only resulted in a thick block of "Others" and the loss of information. Displaying only the label when the company has three or more connections to news pages led also to information loss and at first glance could have led to the assumption that fewer third party companies are connected.

To include all the names, they are interpolated to different shades of grey, just like the connections. To increase readability, the font size of the labels are scaled according to the number of connections. The result looks like a word cloud (see Figure 5.10), where companies with more connections are highlighted, as they are also found on more news pages.

For companies that only establish a few connections to news sites, it is now visible that they exist, but they are not always clearly readable. To solve this problem the label will be highlighted, if there is for example a mouse-over. The highlighted label is written bigger and coloured in black, other labels are slightly faded out, to make it easier to read.

In order to simplify the update pattern, and to prevent too many connections from crossing over during the rearrangement, the third party companies were ordered alphabetically.

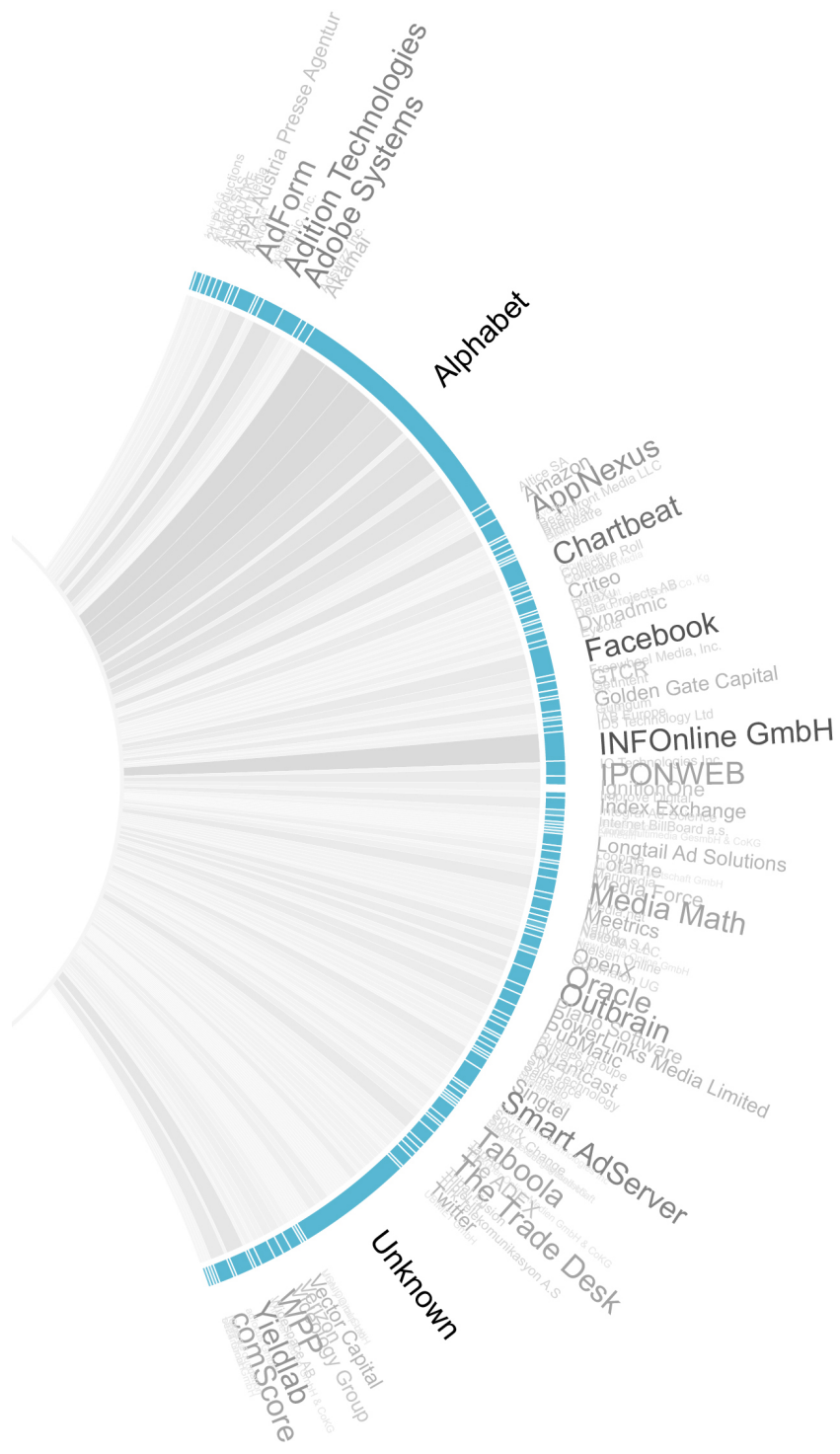


Figure 5.10: The labels of the third parties, highlighting companies with more connections.

## 5.5 Elements

The elements can be sorted in different ways, either all grouped elements and their occurrence are displayed or the elements are sorted per domain by type or country of origin. To distinguish between first and third party elements, third parties are filled with the corresponding type colour and elements originating from the news pages have only a border colour and are not filled.

By default all elements are displayed as a dot inside the circle in the centre, an example can be seen in Figure 5.11, showing the elements of case 1. The colour of the elements depends on the type and the size indicates how often the element was called in relation to the other elements. The position of the elements depends on their occurrence, in the middle of the circle are those that have been called the most often and at the border are those that have occurred the least.

The information is always scaled to the circle in the middle of the visualization, so it can happen that a circle in case 3 is smaller than in case 1, although the element is called more often. This is due to the reason that more elements are loaded in case 3, but they are scaled to the same area, thus the dots appear smaller.

In this case the most occurring element was a pixel that was called 22 times. This value is the grouped one, which was described in section 4.2.

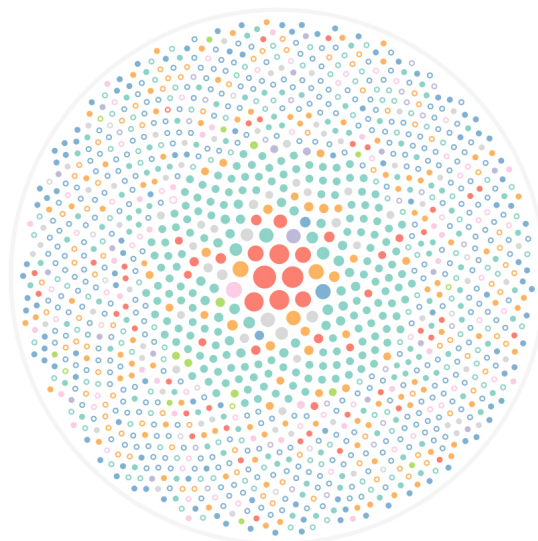


Figure 5.11: All loaded elements are displayed, coloured by type and sized by occurrence.

To find out which domain loads how many element types, the elements can also be sorted by type. An example is shown in Figure 5.12, where the elements from Figure 5.11 are rearranged. The individual points now represent domains and are grouped according to type.

With this display one can determine which element types are loaded most frequently and which domains load particularly many elements of this type. It can be seen that media elements were most often loaded directly from news sites, but a large number of cookies were already tried to be set by third parties, although no interaction took place in case 1, and therefore no cookies were accepted.

The types are also shown in relation to the circle, the biggest dot in the cookies section corresponds to 59 cookies that were tried to be set by this domain.

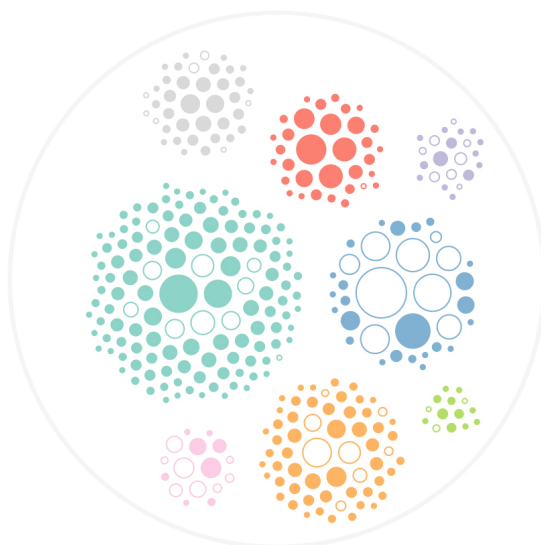


Figure 5.12: Grouping the elements by type and domain.

In order to give an overview from which countries the elements come and in which countries data about the user is potentially processed and collected, the elements can also be sorted by domain and country. The sorting of the elements of case 1 can be seen in Figure 5.13.

Each country is represented by a circle and labelled with its country code. Within the circle are the domains where the owner is from that country. The size indicates the total number of elements that have been loaded, and the colour depends on the type that this domain has loaded most often.

In this case, too, the elements within the circle are shown in relation to the circle. Most domains are from the United States, with 99 domains and 700 grouped elements. The largest domain loading 59 elements and has the most loaded type being cookies. Most elements come from Austrian companies; 762 elements are loaded from 21 domains.

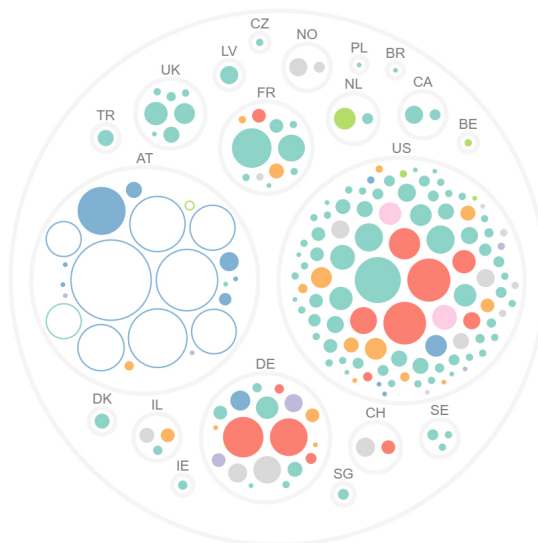


Figure 5.13: Grouping the elements by country and domain.

## 5.6 Interactions

The user can interact with any element of the visualization and additional information is displayed. On mouse over of a news site, besides the domain, the owner, the number of third party connections, as well as how many first and third party elements are loaded is displayed above the main circle in the centre. For the third party companies, the owners additional information about their sub domains are shown, and for the domains their lineage and the number of elements as well as their occurrences.

To gain a better overview of the data it is possible to filter the displayed elements, by selecting a news site and/or a third party company. In Figure 5.14 the news page *oe24.at* is selected and the pixel dot in the centre is hovered, showing the corresponding name, origin and occurrences. After selecting a news site, it is also possible to mouse over another side to highlight their shared elements for better comparison. In Figure 5.15 all elements from the domains of *Facebook* are shown, after selecting *oe24.at* in case 3, accepting the default settings and reading three articles, and showing the most occurring element is *Facebook's* tracking pixel, which is called 21 times. When mousing over this element, all news pages where the same element was found are highlighted.

After sorting by type a domain with a high number of cookies stood out, upon selecting the dot, the data is filtered to show only elements from the domain owner. To further investigate instead of grouping by type, the data is filtered by element to show the individual cookies and on which news sites these can be found on, as seen in Figure 5.16. Furthermore, it is possible to highlight and select all companies from a country, as seen in Figure 5.17, to analyse where the countries are located that may have collected personal data and which elements are loaded by them.

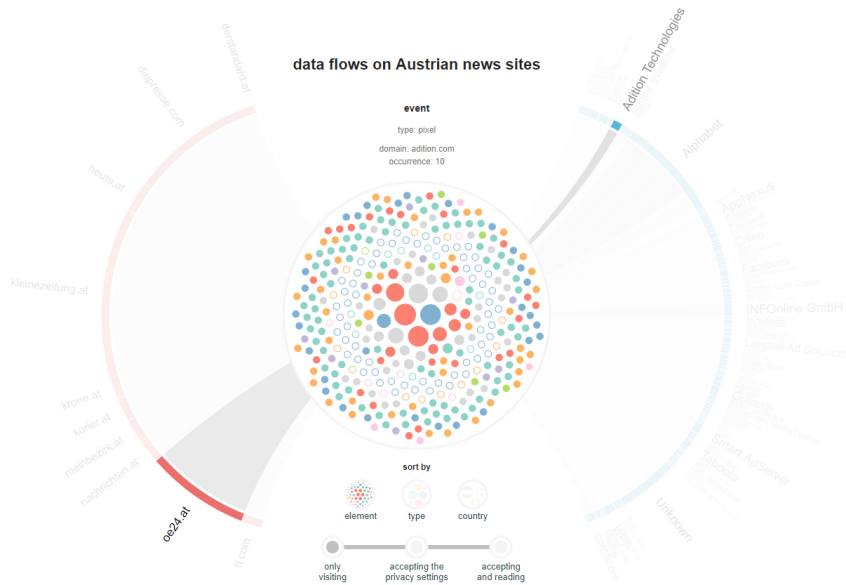


Figure 5.14: Selecting *oe24.at* and highlighting the origin of an element through mouse over.

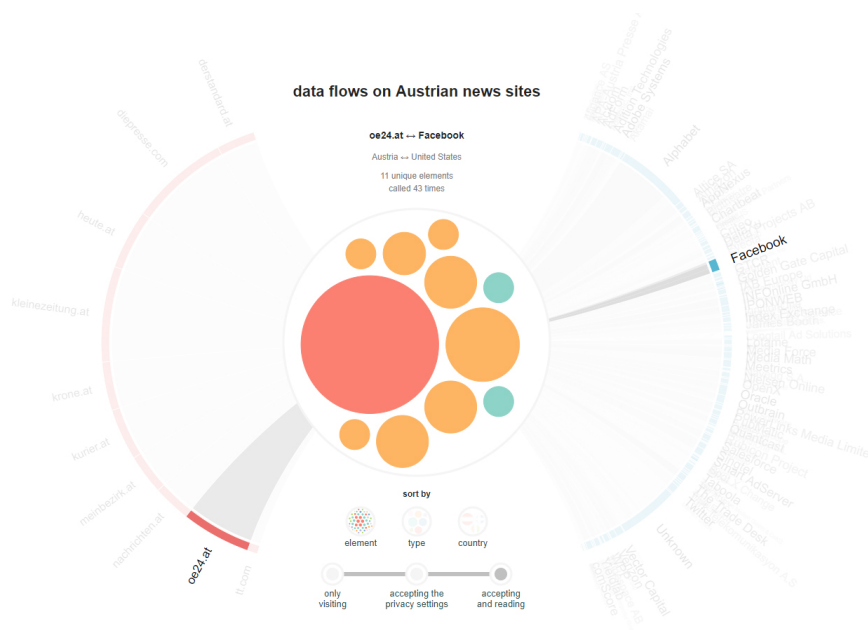


Figure 5.15: Showing which elements from *Facebook* are loaded after visiting *oe24.at* in case 3.

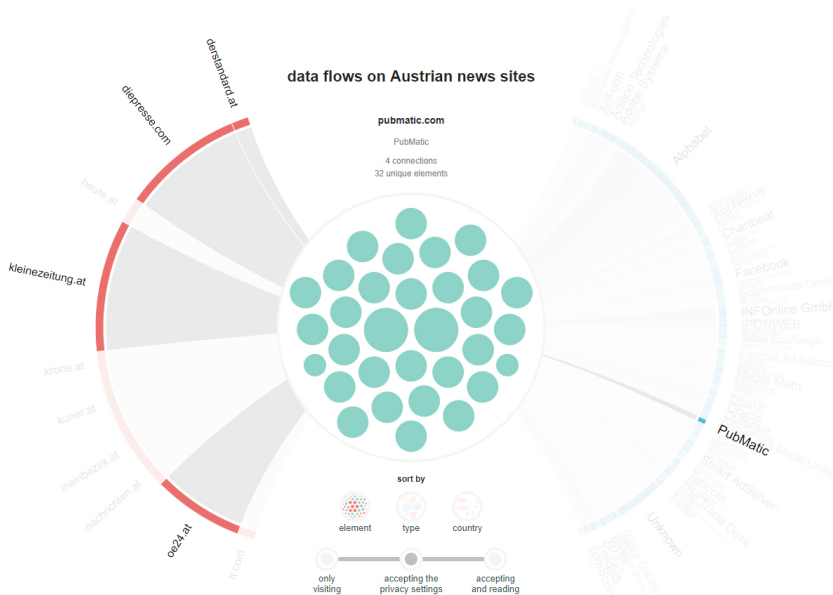


Figure 5.16: Selecting *PubMatic*, showing that only cookies are set by this domain, as well as on which news sites.

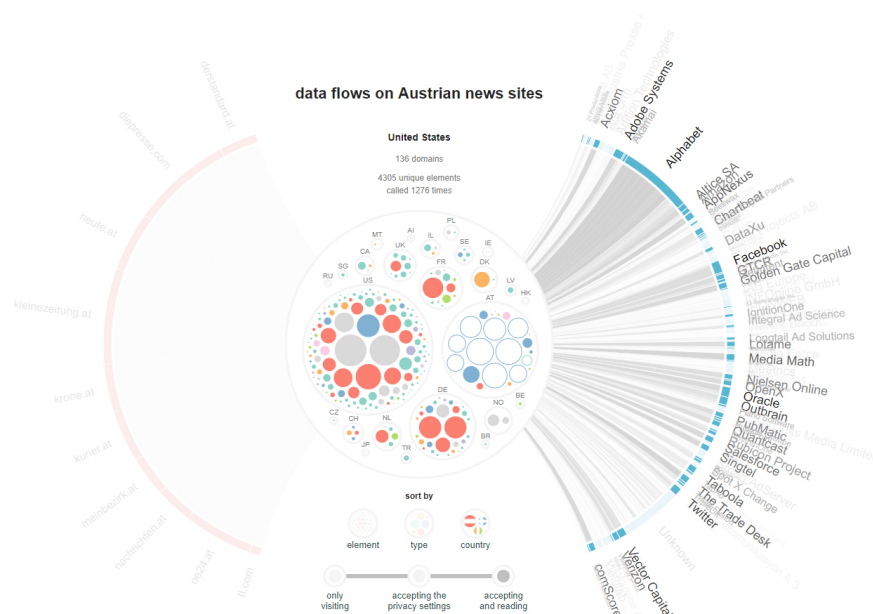


Figure 5.17: On sorting by country the companies from the *United States* are highlighted on mouse over.



## 5.7 Evaluation

The evaluation was conducted in two steps. First, an expert was questioned about the implementation and possible design improvements. Then interviews were conducted about reading the news and advertising in general, as well as testing the visualization, about intuitiveness and the potential to increase the awareness on hidden data flows and possible tracking.

### 5.7.1 Expert Discussion

The visualization was discussed with an expert of information visualization and interaction design to find improvements and if the implementation is intuitive. The topic was explained, as well as the aim of the visualization and the earlier prototypes were shown. The discussion took one hour and the following points were discussed:

1. Can the visualization communicate the information?
2. Is the placement of the layout elements intuitive?
3. Is the navigation clear and easy to understand?
4. Are there difficulties with the interaction of the graph?
5. How can the visualization be improved?

The flow of the data is clear and the different groupings allow the data to be viewed from different perspectives. Since the additional information, that is displayed when interacting with the elements, is always positioned at the same place a quick information acquisition is possible.

One observed problem was the interaction with the elements. In the visualization, the dots can be clicked on for each grouping option (element, type or country), and only those elements that belong to the owner of this element are selected. For example, if a pixel from *google-analytics.com* is clicked on, the elements are filtered so that only those from *Google* are displayed. This can lead to misunderstandings, because the user may assume that when one pixel is clicked, all pixels are selected. To solve this problem, more information about the interaction possibilities should be made available to the user at the beginning, as well as the possibility to implement the legend interactively was suggested. This would allow the user to get a better overview by selecting a certain type and would be more intuitive. By making the legend interactive a pre-existing problem about how to integrate a checkbox into the navigation, that allowed the filtering of the data to only show third party elements and thus reducing the number of elements to allow a better overview, was also solved. As the legend contains two circles, of a first and third party element, the filtering can be done through these elements, allowing continuity in the

legend, that all elements can be interacted with, and the reduction of another navigation element.

A disadvantage of the visualization is that no quantitative readings are possible through the display with the dots. As these are always scaled to the circle in the middle, the user can assume that a smaller dot means a smaller number of occurrences. Since there are 3 times as many grouped elements in data set 3 (accepting and reading 3 articles) as in data set 1 (no interaction), elements with higher occurrence appear smaller in set 3 than elements with lower occurrence in set 1, since more elements have to share the same space. An extreme case is that by filtering the data only one element is displayed, that is then scaled to the whole circle and has the effect that it appears to be called very often. A suggestion for improvement here was to set a maximum size for the dots to prevent at least scaling to the size of the middle circle. Because of the large amount of data, it would not be useful to write the number of calls into the dots, as this would not be readable within the smallest dots.

With the additional representation of the loaded elements, the visualization fulfils its purpose, but due to the many elements and interaction possibilities, the users will need an introduction in order to find their way around optimally. However, the visualization has the potential to shock the users with the sheer amount of data and to make them think about the hidden data flows and possible tracking.

Further improvements:

1. add the legend as a further interaction element to allow sorting by type, for example only showing cookie elements and third parties that set cookies
2. add additional information for each type, explaining the purpose and possible tracking usage
3. colour the connections for the news pages to the centre in the same gradient as the third party domain connections for consistency

The final improvement of the visualization can be seen in Figure 5.18, showing the collected data of case 3 after accepting the default privacy settings, reading three articles and sorting by country. The connections are now coloured in different shades of grey depending on the number of connections, whereby darker grey tones indicate a higher number of connections to third party companies. The additional functionality of filtering the data by type can be seen in Figure 5.19 for case 1, highlighting all domains that set cookies.

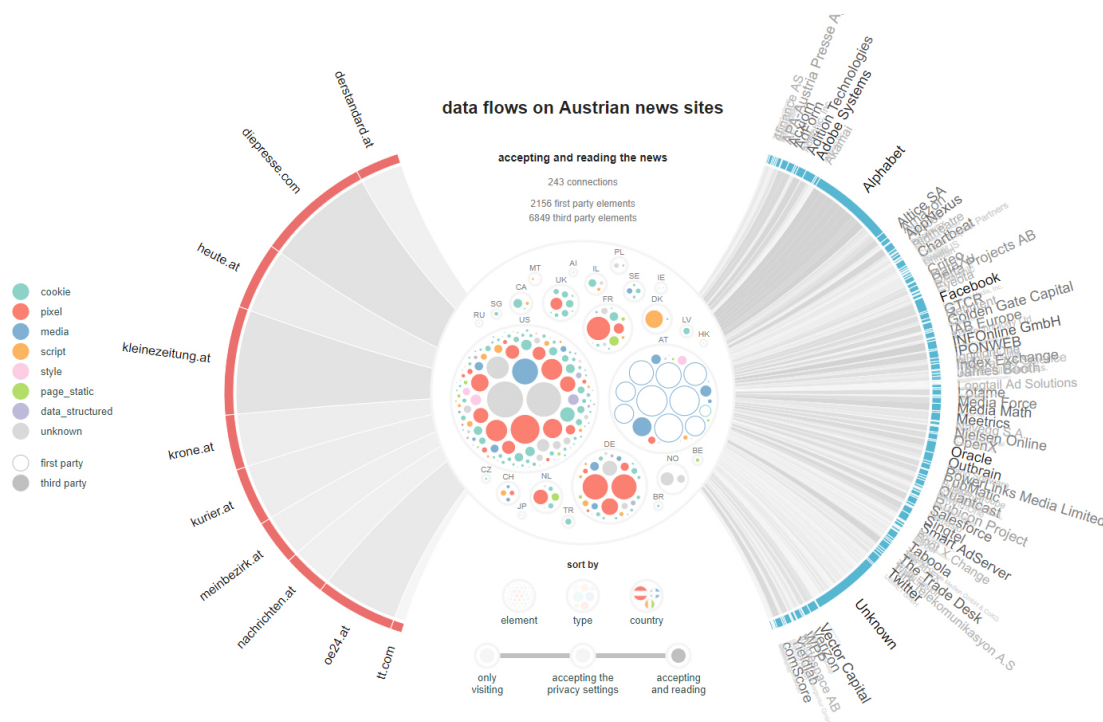


Figure 5.18: Final Visualization after applying the improvements.

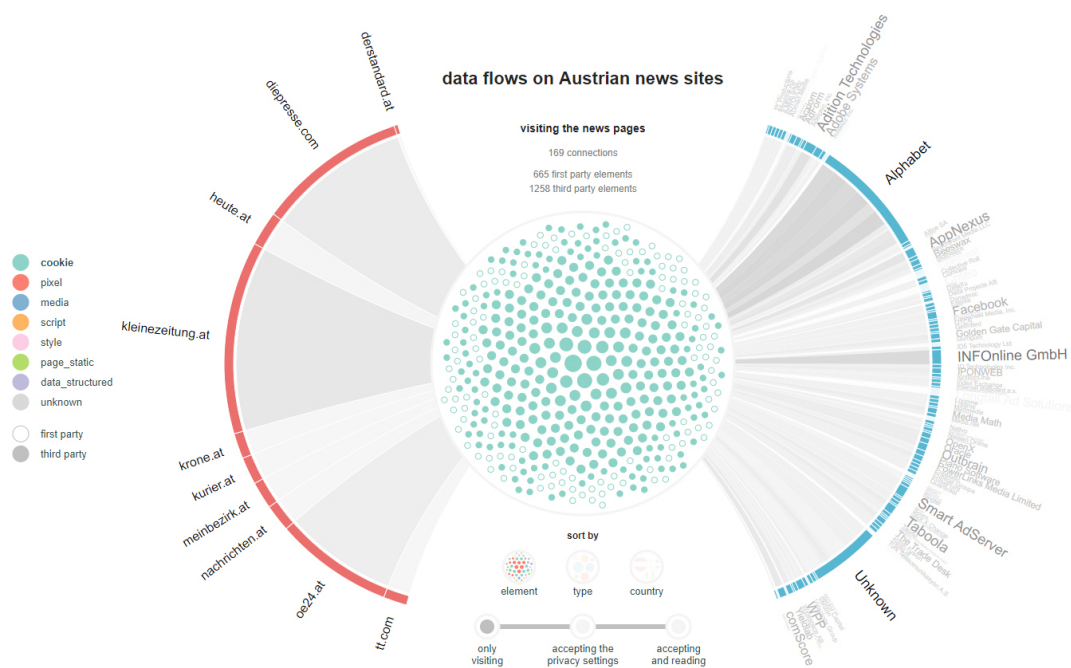


Figure 5.19: Final Visualization after sorting by element type cookie.

### 5.7.2 User Interviews

To test the goal of the visualization, to increase awareness on the topic, as well as the intuitiveness and information density shown, three potential users were interviewed with open questions, to gain general insights on their news consumption, how they handle privacy notifications and online advertisements. The duration was between 40 and 60 minutes, the answers were documented and the users had the possibility to explore the visualization on their own and were encouraged to share their thoughts on their actions. Then the users were asked if there was anything they did not understand and possible ideas for improvement.

How news are consumed was different for all three participants. One interviewee read on three different news sites online and, if available, one newspaper on weekends. The other had a fixed order of which news pages were accessed online. Starting with *orf.at* to find out the most important news and followed by up to four news pages, depending on the time available, that were accessed during the commute to read interesting articles and comments. Both of them rejected suggested contents, on the one hand in order not to be imposed and to inform oneself more consciously and on the other hand because they are “*mostly annoying clickbaits*”. One tried to avoid traces by using different browsers and search engines, and for example not always logging in on *Google*. The third interviewee read rarely and listened rather to the radio news. However, the news feed from the browser start page is often reviewed, where current news from different news sites are suggested and which are read if interested. Whether the articles proposed are based on content already read could not be answered, but if that were the case, the interviewee would not have a problem with it.

The privacy notifications were handled differently. One interviewee stated to read through manageable notices and reject everything that is possible and visible in the main display, but would rather click “accept” if no other option is available. If the notice or the privacy settings are not compact, the user will leave the site because “*I do not want to read 2 hours concentrated*” to understand the purposes and would in this case look for another site without tracking. The second interviewee looks closely and also finds the hidden settings and switches off everything that is possible. But if there are no options available, the privacy notification will be accepted if it is a news page with interesting articles, but on other sites an alternative is searched. The third interviewee stated that the privacy notifications are always accepted as they are, without adjustments. At the introduction of the notices a few were read through, but since then they are always agreed on.

The interviewees agreed that some sites show too much advertising, which they usually ignore. None of them said they clicked on the ads, but they all wondered why this particular advertisement was shown to them. The first interviewee confessed to think a lot about why this specific advertisement was shown and tries to find the connection. Another felt especially haunted by *Amazon*, because after viewing a product, a lot of *Google* ads about the products are shown and wants to know how that works and if personal data is sold from *Amazon* to other companies, “*which is bothersome*”. The third

interviewee also feels disturbed by personalised advertising and tries to ignore it, but also feels followed by products and wants to know why they are shown. The participant had the feeling that “*companies know too much*”, but has no idea how to stop it.

All three interviewees behaved differently online regarding their privacy. One used an ad blocker to protect against tracking, while another automatically deleted all cookies after closing the browser, but had no idea that there are other tracking possibilities besides cookies and wanted to know more about them. In general, they all had no problem with advertising on websites as long as they are not too much, too intrusive or too personalized. As an example of “*too much*” *heute.at* was given, that “*can not be viewed without an ad blocker*”, because of the amount of intrusive advertisements. The interviewee with the advertising blocker stated that this is why advertising is blocked on all pages, because individual settings take too much effort. The participant also always wondered why *derstandard.at* insists on deactivating the ad blocker to see the content and was not aware that there is an option to pay. None of the interviewees would pay to view ad free content. One stated, that a subscription model for several news sites as a package would be considered, if there were such an option, but to pay for each side individually costs too much.

The testers were at first a little overwhelmed with the information shown in the visualization and needed a little time in order to find their way around. One was a little hesitant to interact with the visualization at first, afraid to do or click on something wrong, whereas the other two would automatically hover over elements that were unclear, for example the meaning of the thickness of each news site, showing the number of third party connections, and found the additional information. In general, all testers were first or even only interested in the news pages, which they come in contact with or are actively reading. One stated to want an option to hide the other news sites whereas the other two had no problem with them. One interviewee stated the other news sites are important, because if an element is found on “*their*” news site and on others then this shows that the elements may follow them. The data was sorted in different ways, whereas each participant had a different sorting preference (elements, types or countries), which they rarely changed and mostly interacted with the different news sites, third parties and the different data sets to observe the changes.

All participants stated that they were not as aware of what was going on in the background and were surprised by the number of connections and that so many cookies were tried to set without interaction. Two asked if there was a possibility to look at the visualization again after the first shock settled, to gain more insights at details and how the news sites are connected. The participant, who always accepts the privacy notification, said that thought the visualization more awareness was gained and to change the behaviour to always accept the default settings. One tester wanted more information about the individual companies, what exactly their field of business is and how they earn their money. Another suggestion for improvement was an intro, which explains to the user step by step how to interact with the visualization, for example, that when one page is selected and another is hovered, that the highlighted elements are the ones shared

by both news sites, because this was not clear to the participant. Furthermore, a FAQ page could be integrated, which answers the most important questions and also provides additional information.

Finally, during the interviews it was observed that in order to reach more people and explain how they can be tracked and how the techniques work, the visualization could be part of a larger project, that also explains what a domain is, how a cookie is set, or how a pixel can be used to track user behaviour. It was observed that through the visualization more awareness was gained by showing the background activities and providing different data sets to observe the changes. However, by interacting with the data the participants asked more questions about tracking, data collection and the different third party companies, that are currently possible to answer with the visualization, but shows a greater interest in the topic that was raised by the visualization.

# Conclusion

In this chapter the insights gained after exploring adTech and collecting data as well as possible privacy risks are discussed. Then this thesis concludes with a summarization of the findings and further improvements.

## 6.1 Discussion

If the user visits a website and decides to disagree to the stated tracking and collection of personal data, but finds no way in the privacy notifications to decline, and therefore leaves the site, it was shown that tracking was already taking place and third parties tried to set or access their cookies to update information about the user. An observed problem was the displaying of advertisement before the user has the chance to accept or decline the usage of tracking technologies to display personalized advertisements. Since most publishers are taking part in the adTech ecosystem and selling space on their website to ad platforms, advertising auction and cookie matching are already taking place and the more information a publisher provides about the user or the adTech companies have already collected, the higher they are bidding for the right to display their advertisement if the user fits the targeted audience. As a result, publishers may not know which third parties the user comes into contact with on their website, but are nonetheless responsibly for all activities on their website.

In general, it has been observed that only by interacting with the news site and reading articles, the extent of cookies and tracking pixels could be illustrated, if the user agrees to the default settings of the privacy notification. The visualization can be used to illustrate how many third parties the user comes into contact with through the use of the news pages. The insights are limited in the way, that only tendencies can be observed and to find out the exact purpose of an element or whether the element is malicious further research and analysis is necessary. Furthermore, to find out which of these companies are specialized in tracking and what exactly the task of each element is,

would be interesting to evaluate in a future work. Through the visualization, collecting and tracking tendencies can be identified that the same element or group of elements was loaded on several websites.

It is possible for third parties to collect analytic or personal data, by placing content on the side or by participating in cookie matching or advertisement auctions to gain access. When a connection is established with the user by default, the IP address of the user is transmitted to allow a successful communication and thus reveal the approximate location and a possible way to identifying the user. An exception is when the users protect themselves, for example, by using the *Tor* browser.

In order to determine exactly which elements are used specifically for tracking, the content would have to be analysed, since some scripts can be important for the functionality of the website. With the help of the visualization, sorting by third party companies, it can be concluded that if a company sets cookies, calls up tracking pixels and executes a script, a tracking tendency can be recognized. Pixels usually occur in combination with scripts, as this method is used by adTech companies to collect data about the user, even if they try to protect themselves and block *JavaScript* from third parties by default.

Except of a pixel from *krone.at*, which is recurring in all three cases, pixels were only used by third party companies. Most domains that use pixels have 1 to 3 main pixels that were always called with different queries, for example a pixel from *Google Analytics* named “*collect*” is called 264 times in case 3 on all 10 news pages, whereas the highest occurrence was on *krone.at* with 68 calls.

Because news pages have been analysed, which generally contain many images and are therefore not easy to distinguish from advertising just by type. In general, the trend has found that images for articles, loaded by first or third parties, usually describe the content in the title, for example “*brexit2017.png*” and images for personalized advertising either consist of a code “*9eafed76a0c.48.jpg*” or were usually loaded using a query, for example: “*activityi;src=6345355;type=kurie0;cat=kurie0;ord=aD5cE7b18B747aec;...*”.

Companies were observed that load many different elements with only queries, which could not be evaluated afterwards. For example, an unknown element is displayed in the visualization that was called up frequently and whose element is for example named “/”, because the third party page called up elements with “*exampleAds.com/?query*”. During the classification these elements were classified as unknown, because different element types were returned or the elements were no longer found. A potential tracking possibility, which many web site publishers are not aware of, is also the inclusion of fonts from third parties. Most style elements were fonts loaded from the *Google APIs* and therefore a potential source of information for *Google*.

The GDPR of the EU requires that personal data is only collected and cookies are only set with the explicit consent of the user, with the exception of necessary cookies. An overview of all third party elements and their type, that were loaded in case 1 per news site, can be seen in Figure 6.1, which are used to discuss if the news page is privacy friendly. Generally, each loaded third party element is a potential source for the third



party domain to collect personal data and to track the user on different sites. In order to determine whether the news sites are GDPR compliant the loaded elements would have to be further investigated, but the use of cookies by third parties could be an indicator of GDPR compliance, as these require explicit consent. In some cases it is possible that third party cookies can be considered of necessary interest to the news site, but would have to be verified individually by the responsible data protection authority.

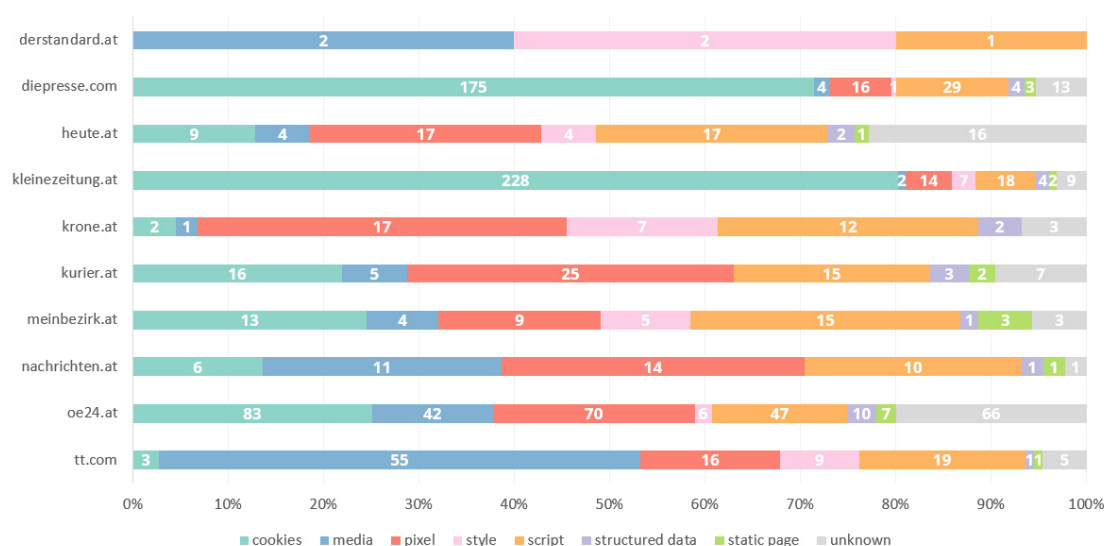


Figure 6.1: Third party elements loaded after visiting with no interaction and grouped by type for each news site.

It is positively notable that the *derstandard.at* only loads 5 elements from another source. These all come from the domain *staticfiles.at*, which also belongs to the *Standard Verlagsgesellschaft*. When visiting *derstandard.at* the user is faced with a privacy wall, and has the option of accepting cookies for analysis and advertising, or to purchase a paid subscription and thus use the website without advertising and cookies, as seen in Figure 4.4. This privacy notification is taken seriously and *derstandard.at* is the only examined news site on which no third party cookies are set on the user's computer by only visiting the site. An advantage of this method is that if the user decides not to agree and leaves the website, no third parties would have had the opportunity to obtain information about the user.

The other nine websites examined are not very privacy friendly by evaluating based on types as well as the number of connections to third parties. In the 45 seconds each news site was examined, a large number of cookies, tracking pixels and scripts were loaded. Strictly necessary cookies, which are essential for a website to function and do not need explicit consent, are often first party cookies, but are also required to be explained to the user “*what they do and why they are necessary*” [43]. Thus for further research the collected data could be compared with the available information that must be provided

by the site to determinate GDPR compliance. However, overall 353 third party cookies were set with not interaction from the user which is an invasion in the users privacy.

A noticeably high number of cookies can be found at *kleinezeitung.at* with 228 third party cookies which make up 80% of all third party elements of this website and are set by 92 different domains. Most of the cookies are from a company in the United States called *PubMatic*, which tries to set 30 different cookies, when the user visits the site. The company *Smart AdServer* from France sets 22 cookies from two domains, followed by *StickyADS.tv* from the United States with 12 cookies. *PubMatic* is found in the section “*Personalization*” in the privacy policy and is one of the 343 listed companies (which are not ordered alphabetically). The purpose is listed as “*Storage and Access to Information*” and the company is used for “*Personalization*”, “*Selection, placement and evaluation of advertisements*”, “*Selection, placement and evaluation of content*” and “*Evaluation*” and thus not strictly necessary. At the website of *diepresse.com* 70 domains try to set 175 cookies, which represents 71% of all third party elements. The companies *PubMatic*, with 29 cookies, and *Smart AdServer*, with 21 cookies, are also represented here. In third place is the British company *ID5 Technology* with 8 cookies.

The sites *kleinezeitung.at* and *diepresse.com* both use the same privacy layout, as seen in the Figures 4.7 and 4.6. They listed 466 and 468 third party companies, that would have access to the collected data after consenting, but both privacy notices indicate under “*Information*” that as long as no cookies are accepted, only the absolutely necessary ones are set to provide the offered services. *kleinezeitung.at* lists under necessary cookies 7 third party companies, amongst others *Google Analytics*, whereas *diepresse.com* only lists two. Since far more third parties tried to set cookies based on the collected data both news sites appear to not be GDPR compliant.

The news site *oe24.at* loads the most third party elements, whereby 83 of the 331 loaded elements are cookies (25%). The cookies are set by 34 different domains, most of them from the American company *Outbrain*, which tries to sets 10 cookies from 2 domains. The privacy notification allows only the option to accept, that “*this website uses cookies*” and in the privacy policy it is noted, that to deny the usage of tracking and marketing cookies the user has the option to opt-out on two sites from the US and EU or to block third party cookies in the browser settings. Furthermore, the usage of the *Facebook Pixel* is described as a legitimate interest as well as the *Facebook Plugin*, but it is mentioned that the news site has no control of the personal data collected by *Facebook* and if a user with a *Facebook* account does not want the company to connect information gathered on this site to his or her profile, the user should log out of their account and delete all cookies before visiting *oe24.at*, making the option rather complex instead of limiting the access at least to the point where the user agrees to this functionality. Overall the website only allows opt-outs through different third party services and all third party elements are seen as legitimate interest or out of their control. Thus the user is exposed to many third party companies, that are not necessary for the site to function just by opening *oe24.at*, which is not privacy friendly.

It is notable that besides *derstandard.at* also *kurier.at* offers a payment abo without

advertising (with the exception of two advertisers), but before the user can decide or log in, cookies are already set by third parties and the highest number of pixel calls in this case is being conducted, indicating the collection of data.

In total, 301 unique cookies were set on the nine news pages from 118 different domains, which were accessed 535 times. Furthermore it was found that 79 domains only set cookies and provide no other elements. These domains have a total of 192 unique cookies which are accessed 330 times, which makes up the majority of the third party cookies. Since a total of 169 unique connections to domains were established, 70% of them set cookies without the user's knowledge.

## 6.2 Summary

This thesis started with the idea to explore the influences of advertising technologies and find a way to encourage awareness on the topic. To gain a better understanding how the user comes in contact with adTech, first the different roles used to distribute advertisement and the methods to collect personal data were researched. The limitations of this work were, that the adTech industry is always in development of new ways to track the user and methods to target and personalize advertisements, and the impact on the public of the usage of these technologies is not always immediately visible. The examples about political influences are already a few years old, but their specific influences as well as the involved players are still not fully resolved, as the involved companies are not cooperative to admit wrongdoing and that the privacy of users has been violated. The introduction of the GDPR of the EU is a step in the right direction to protect the privacy of the users online. However, since personal data is considered an economic value, it was explored that the implementation is often only casually or the users are nudged to accept the collection and the analysis of their data, to increase profit.

With the insights gained after exploring the sphere of adTech as well as already existing projects to increase the users online privacy, Austrian news sites were selected to encourage more awareness about adTech. On one hand the news pages are a daily source of information for many people and on the other hand it was shown that tracking and third party content can be found more often on these sites. For this purpose data about third party connections and loaded elements were collected on the 10 most popular news sites of Austria. To gain more insights the data was collected for three different cases and further analysed. With the help of the tool *webXray*, to collect the data, the news sites were visited with no interaction, with only accepting the default privacy settings and to simulate a user the default privacy settings were accepted as well as three articles opened. The cases were chosen to increase the awareness about how personal data can be collected and how important it is to take advantage of the option to control how personal data is used for different purposes which must be provided by the publishers to comply to the GDPR. To gain better insights on the loaded elements, the elements were classified in different groups: cookie, media, pixel, style, script, structured data, static page as well as unknown if the element could not be assigned a type. Furthermore, to show

connections between domains and their owning companies, the owners as well as their company lineage was researched. This allowed grouping all domains of one company, for example *Alphabet (Google)* to gain insight into on how many news sites the user comes into contact with the company. In this step domains that could not easily be assigned to a company were observed, because their information was either redacted for privacy or they were using another third party company in their stead as the domain registrant.

To make the collected data more accessible to the user, the design process as well as the implemented prototypes were described which lead to the final interactive visualization. To understand if the visualization can be used to increase the awareness on adTech two evaluations were conducted. First an expert was questioned about further improvements and the intuitiveness of the design. It was found that the interactive visualization can be used to show the hidden data flows and by filtering the data allows the user a detailed insight into the data and possible tracking. A possible disadvantage is the amount of data displayed, as well as the scaling of the data to increase the visibility of the elements, thus no quantitative readings are possible without interacting with the visualization. To test the final design potential users were interviewed. The aim was to first understand the users thoughts on online privacy and their news consumptions with open questions. The second step was the interaction with the visualization to further understand the design and what information the users can gain after interacting with the data. Finally the insights gained were discussed with the users and all stated that they were afterwards more aware of the background activities and the possible privacy risks and wanted to change their online behaviour, for example not always accepting the default privacy settings. Furthermore, it was observed that the users asked more questions about the topic after interacting with the visualization, which on one hand shows that the visualization raised more interest in the topic, and on the other hand the potential to further increase the design.

### 6.3 Future Work

In a future work the collected data could be further analysed to find out what functionality each loaded element has, for example, whether it is essential for the presentation and use of the website, or whether it serves another purpose, such as analytic, performance or tracking the behaviour of the user. This additional data can be included in the visualization to allow more detailed insight which elements invade the users privacy. Furthermore, during the interviews it was concluded that for not so technologically experienced people, additional information would help to understand the tracking ecosystem, for example how cookies are set and how a pixel can be used to track online behaviour. The visualization increases the awareness on the topic, thus the users asked more questions after interacting with the data about how the adTech industry works, then the visualization currently can answer. Therefore, it is possible in a future work, that the visualization could become part of an educational page where additional information about advertising techniques are described, which allows the users to get even deeper insights into the adTech ecosystem.

# List of Figures

2.1	The interactions between the players in the adTech ecosystem, from [2]. . . . .	6
3.1	Leave EU: Sending £350 mio. a week to the EU, [22]. . . . .	16
3.2	Leave EU: £350 mio. to the National Health Service (NHS) instead of the EU, [23]. . . . .	16
3.3	Leave EU: £350 mio. to the EU or to Yorkshire, [23]. . . . .	16
3.4	Leave EU: British youth unemployment compared to other EU countries, [23].	17
3.5	Leave EU: Turkey joining the EU, [23]. . . . .	18
3.6	Leave EU: Visa-free travel for Turkish citizens and new border of the EU besides Syria and Iraq, [23]. . . . .	18
3.7	Leave EU: EU blocks ability to protect polar bears, [23]. . . . .	19
3.8	Leave EU: EU wants to ban tea kettles, [23]. . . . .	19
3.9	Ad Library: <i>Facebook</i> pages with the highest ad spending in Austria. . . . .	23
3.10	Ad Library: Two examples of advertisement from <i>Greenpeace</i> . . . . .	24
3.11	Ad Library: information about the ad as well as demographics of the targeted audience. . . . .	25
3.12	A simple cookie notice, stating how first party cookies are used, with one option to accept the usage . . . . .	27
3.13	This privacy notification accepts closing, dismissal, clicking on links or buttons, continuing browsing and accepting as acceptance . . . . .	27
3.14	This simple cookie notice leads to advanced settings. . . . .	28
3.15	The overview of the advanced cookie settings blocked by advertisement. . . . .	28
3.16	Third parties can be toggled to deny consent or require individual opt-out.	29
3.17	A cookie notice highlighting the “ <i>Accept all and save</i> ” button to nudge the user to accept and hiding “ <i>Save choices</i> ”. . . . .	30
3.18	A cookie notice nudging the user to click “ <i>select all</i> ” to accept the usage of cookies. . . . .	31
3.19	On the left is the cookie notice with “ <i>Analytics Cookies</i> ” selected, on the right the option is disabled, leading to the appearance of a new button “ <i>Allow All</i> ” in place of “ <i>Save Settings</i> ”. . . . .	31
3.20	This privacy notification can only be closed by choosing “ <i>Accept all</i> ” or by selecting one additional option to “ <i>Save preferences</i> ”. . . . .	32
3.21	Dark pattern strategies, from [48]. . . . .	34
		87

3.22	The Firefox extension <i>Lightbeam</i> after visiting 6 websites. . . . .	36
3.23	Austria's top 50 websites: third party cookies and their connections with no user interaction visualized with <i>Lightbeam</i> . . . . .	38
3.24	Overview of <i>Trackography</i> showing all available countries. . . . .	39
3.25	<i>Trackography</i> illustrates where data travels after visiting <i>meinbezirk.at</i> . . .	39
3.26	The browser extension <i>Privacy Badger</i> on the news site <i>newyorker.com</i> . . .	41
3.27	The browser extension <i>Ghostery</i> on the news site <i>kleinezeitung.at</i> . . . . .	42
4.1	Top 10 of the most popular online news sites in Austria from September 2019, from [61]. . . . .	46
4.2	Example of opening a tracking pixel. . . . .	52
4.3	A privacy protection company is listed instead of the owner information of <i>cleverpush.com</i> . . . . .	53
4.4	Pay wall of <i>derstandard.at</i> offering to accept cookies or pay to view the content ad free. . . . .	55
4.5	The privacy notification of <i>oe24.at</i> nudges the user to accept third party cookies. . . . .	56
4.6	The privacy settings of the privacy notification of <i>kleinezeitung.at</i> . . . . .	56
4.7	The privacy settings of the privacy notification of <i>diepresse.com</i> . . . . .	57
4.8	Advertisement in relation to content on <i>heute.at</i> . . . . .	57
4.9	Sponsored content from <i>diepresse.com</i> , including two advertisements and one news article. . . . .	58
4.10	Sponsored content from <i>meinbezirk.at</i> , including three news articles (top row) and three advertisements (bottom row). . . . .	58
4.11	Sponsored content from <i>oe24.at</i> , including one advertisement and one news article. . . . .	58
5.1	Concept of a Sankey visualization showing third party connections and third party lineage information. . . . .	60
5.2	Prototype implementation of a Sankey diagram showing the flow of third party connections for each news page. . . . .	61
5.3	Section of the Sankey, showing the highlighting of the third party company <i>DoubleClick</i> . . . . .	62
5.4	Comparison between the default pattern (left) and phyllotaxis pattern (right). . . . .	62
5.5	Circle packing prototype showing the number of third party connections for each news page. . . . .	63
5.6	Prototype of showing all loaded elements inside a circle for each case, coloured by type (top) or country with only third party elements (bottom). . . . .	64
5.7	First sketches of the final visualization. . . . .	65
5.8	First working Prototype showing the connections of the news sites on the left, all loaded elements in the middle circle and the third parties on the right. . . . .	66
5.9	Colour gradient for the third party connections. . . . .	68
5.10	The labels of the third parties, highlighting companies with more connections. . . . .	69
5.11	All loaded elements are displayed, coloured by type and sized by occurrence. . . . .	70
88		

5.12	Grouping the elements by type and domain. . . . .	71
5.13	Grouping the elements by country and domain. . . . .	72
5.14	Selecting <i>oe24.at</i> and highlighting the origin of an element through mouse over. . . . .	73
5.15	Showing which elements from <i>Facebook</i> are loaded after visiting <i>oe24.at</i> in case 3. . . . .	73
5.16	Selecting <i>PubMatic</i> , showing that only cookies are set by this domain, as well as on which news sites. . . . .	74
5.17	On sorting by country the companies from the <i>United States</i> are highlighted on mouse over. . . . .	74
5.18	Final Visualization after applying the improvements. . . . .	77
5.19	Final Visualization after sorting by element type cookie. . . . .	77
6.1	Third party elements loaded after visiting with no interaction and grouped by type for each news site. . . . .	83



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.  
The approved original version of this thesis is available in print at TU Wien Bibliothek.



# List of Tables

2.1	Difference between HTTP Cookie, Flash Cookie and HTML5 LocalStorage [2, 13] . . . . .	9
4.1	Selected news sites with their domain names. . . . .	47
4.2	Comparison of third party connections for each case and domain. . . . .	48
4.3	Overview of total and unique connections as well as first and third party elements per case. . . . .	49
4.4	The total amount of connections to third parties per news site and case. . . . .	49
4.5	Overview of total and grouped elements per case. . . . .	51
4.6	Top 10 third party owners and their number of occurrences on the 10 observed news pages per case. . . . .	54
4.7	Overview if the news sites offer privacy adjustments or payment for ad free content. . . . .	55
5.1	Element types and their assigned colour. . . . .	67
5.2	Colour scheme for arc elements. . . . .	68
1	Classified Element Types of <i>derstandard.at</i> . . . . .	93
2	Classified Element Types of <i>diepresse.com</i> . . . . .	94
3	Classified Element Types of <i>heute.at</i> . . . . .	94
4	Classified Element Types of <i>kleinezeitung.at</i> . . . . .	95
5	Classified Element Types of <i>krone.at</i> . . . . .	95
6	Classified Element Types of <i>kurier.at</i> . . . . .	96
7	Classified Element Types of <i>meinbezirk.at</i> . . . . .	96
8	Classified Element Types of <i>nachrichten.at</i> . . . . .	97
9	Classified Element Types of <i>oe24.at</i> . . . . .	97
10	Classified Element Types of <i>tt.com</i> . . . . .	98
11	Top domains with lineage and news page occurrence of case 1. . . . .	99
12	Top domains with lineage and news page occurrence of case 2. . . . .	100
13	Top domains with lineage and news page occurrence of case 3. . . . .	101



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.  
The approved original version of this thesis is available in print at TU Wien Bibliothek.

# Appendix

## Classified Element Types

<i>derstandard.at</i>	Case 1		Case 2		Case 3	
	1 <sup>st</sup> party	3 <sup>rd</sup> party	1 <sup>st</sup> party	3 <sup>rd</sup> party	1 <sup>st</sup> party	3 <sup>rd</sup> party
cookie	1		17	10	17	72
media		2	140	7	157	22
pixel				10		122
style		2		16		20
script		1	1	20	8	58
structured data				2		6
static page	1		1	3	1	12
unknown			6	3	16	56

Table 1: Classified Element Types of *derstandard.at*

<i>diepresse.com</i>	Case 1		Case 2		Case 3	
	1 <sup>st</sup> party	3 <sup>rd</sup> party	1 <sup>st</sup> party	3 <sup>rd</sup> party	1 <sup>st</sup> party	3 <sup>rd</sup> party
cookie	15	175	30	195	30	255
media	47	4	84	4	98	50
pixel		16		52		287
style	11	1	11	1	11	3
script	2	29	2	44	2	123
structured data	4	4	5	6	5	25
static page		3		5		21
unknown	1	13	0	17	2	168

Table 2: Classified Element Types of *diepresse.com*

<i>heute.at</i>	Case 1		Case 2		Case 3	
	1 <sup>st</sup> party	3 <sup>rd</sup> party	1 <sup>st</sup> party	3 <sup>rd</sup> party	1 <sup>st</sup> party	3 <sup>rd</sup> party
cookie	8	9	10	19	13	149
media	35	4	262	15	289	170
pixel	0	17		18		132
style	2	4	2	4	4	20
script	4	17	8	17	16	69
structured data	2	2	2	2	2	38
static page	1	1	1	1	1	14
unknown		16	4	24	12	230

Table 3: Classified Element Types of *heute.at*

<i>kleinezeitung.at</i>	Case 1		Case 2		Case 3	
	1 <sup>st</sup> party	3 <sup>rd</sup> party	1 <sup>st</sup> party	3 <sup>rd</sup> party	1 <sup>st</sup> party	3 <sup>rd</sup> party
cookie	14	228	30	204	38	233
media	25	2	143	2	174	34
pixel		14		103		470
style	2	7	2	7	2	10
script	8	18	8	43	11	145
structured data	4	4	5	4	10	14
static page	0	2		8		27
unknown	3	9	3	18	10	183

Table 4: Classified Element Types of *kleinezeitung.at*

<i>krone.at</i>	Case 1		Case 2		Case 3	
	1 <sup>st</sup> party	3 <sup>rd</sup> party	1 <sup>st</sup> party	3 <sup>rd</sup> party	1 <sup>st</sup> party	3 <sup>rd</sup> party
cookie	14	2	19	6	22	111
media	109	1	202	2	256	21
pixel	1	17	1	26	1	130
style	17	7	17	7	26	12
script	20	12	20	24	28	31
structured data	4	2	4	7	10	14
static page				1		4
unknown	4	3	4	8	13	43

Table 5: Classified Element Types of *krone.at*

<i>kurier.at</i>	Case 1		Case 2		Case 3	
	1 <sup>st</sup> party	3 <sup>rd</sup> party	1 <sup>st</sup> party	3 <sup>rd</sup> party	1 <sup>st</sup> party	3 <sup>rd</sup> party
cookie	21	16	23	84	23	134
media	25	5	50	8	95	139
pixel		25		44		166
style	3		3	2	5	12
script	3	15	3	20	4	47
structured data	1	3	1	4	14	18
static page		2		5		11
unknown		7		43		229

Table 6: Classified Element Types of *kurier.at*

<i>meinbezirk.at</i>	Case 1		Case 2		Case 3	
	1 <sup>st</sup> party	3 <sup>rd</sup> party	1 <sup>st</sup> party	3 <sup>rd</sup> party	1 <sup>st</sup> party	3 <sup>rd</sup> party
cookie	7	13	9	24	9	87
media	17	4	88	5	118	32
pixel		9		43		190
style	3	5	3	5	5	7
script	3	15	5	30	14	78
structured data	1	1	1	2	1	2
static page		3		6		14
unknown	1	3	1	11	3	118

Table 7: Classified Element Types of *meinbezirk.at*

<i>nachrichten.at</i>	Case 1		Case 2		Case 3	
	1 <sup>st</sup> party	3 <sup>rd</sup> party	1 <sup>st</sup> party	3 <sup>rd</sup> party	1 <sup>st</sup> party	3 <sup>rd</sup> party
cookie	11	6	18	12	20	72
media	35	11	40	14	161	45
pixel		14		32		195
style	11		31		32	
script	35	10	36	36	41	63
structured data	6	1	11	2	51	6
static page		1		8		15
unknown	1	1	1	18	1	115

Table 8: Classified Element Types of *nachrichten.at*

<i>oe24.at</i>	Case 1		Case 2		Case 3	
	1 <sup>st</sup> party	3 <sup>rd</sup> party	1 <sup>st</sup> party	3 <sup>rd</sup> party	1 <sup>st</sup> party	3 <sup>rd</sup> party
cookie	9	83	15	95	20	146
media	59	42	97	51	175	206
pixel		70		89		305
style	7	6	8	9	9	22
script	12	47	12	50	17	126
structured data		10		10		16
static page	2	7	2	10	2	16
unknown	1	66	2	43	2	222

Table 9: Classified Element Types of *oe24.at*

<i>tt.com</i>	Case 1		Case 2		Case 3	
	1 <sup>st</sup> party	3 <sup>rd</sup> party	1 <sup>st</sup> party	3 <sup>rd</sup> party	1 <sup>st</sup> party	3 <sup>rd</sup> party
cookie	24	3	27	3	27	3
media	5	55	8	137	9	152
pixel		16		23		188
style		9		7		8
script		19		20		20
structured data	2	1	2	1	11	1
static page		1		1		1
unknown	1	5	1	7	2	19

Table 10: Classified Element Types of *tt.com*



## Top Domains per Case

Case 1 - no interaction		
Occurrence	Domain	Lineage
9	doubleclick.net	DoubleClick > Google > Alphabet
9	google-analytics.com	Google Analytics > Google > Alphabet
9	googletagmanager.com	Google Tag Manager > Google > Alphabet
8	google.com	Google > Alphabet
8	google.at	Google > Alphabet
8	iocnt.net	INFOnline GmbH
6	adition.com	Adition Technologies
6	gstatic.com	Google > Alphabet
5	adnxs.com	AppNexus
5	googleapis.com	Google APIs > Google > Alphabet
4	scorecardresearch.com	ScorecardResearch > comScore
4	googletagservices.com	Google Tag Manager > Google > Alphabet
3	taboola.com	Taboola
3	adform.net	AdForm
3	bidswitch.net	Bidswitch > IPONWEB

Table 11: Top domains with lineage and news page occurrence of case 1.

Case 2 - accepting the privacy notifications		
Occurrence	Domain	Lineage
10	doubleclick.net	DoubleClick > Google > Alphabet
10	google-analytics.com	Google Analytics > Google > Alphabet
10	googletagmanager.com	Google Tag Manager > Google > Alphabet
10	iocnt.net	INFOonline GmbH
8	google.com	Google > Alphabet
8	google.at	Google > Alphabet
8	googlesyndication.com	AdSense > Google > Alphabet
7	adition.com	Addition Technologies
7	gstatic.com	Google > Alphabet
7	googletagservices.com	Google Tag Manager > Google > Alphabet
6	adform.net	AdForm
6	adnxs.com	AppNexus
6	yieldlab.net	Yieldlab
6	googleapis.com	Google APIs > Google > Alphabet
5	scorecardresearch.com	ScorecardResearch > comScore

Table 12: Top domains with lineage and news page occurrence of case 2.

Case 3 - accepting the privacy notifications and reading three articles		
Occurrence	Domain	Lineage
10	doubleclick.net	DoubleClick > Google > Alphabet
10	google-analytics.com	Google Analytics > Google > Alphabet
10	googletagmanager.com	Google Tag Manager > Google > Alphabet
10	iocnt.net	INFOnline GmbH
10	googlesyndication.com	AdSense > Google > Alphabet
10	google.at	Google > Alphabet
10	google.com	Google > Alphabet
10	googletagservices.com	Google Tag Manager > Google > Alphabet
9	adnxs.com	AppNexus
9	adition.com	Adition Technologies
9	bidswitch.net	Bidswitch > IPONWEB
9	mathtag.com	Media Math
9	adsrvr.org	The Trade Desk
9	gstatic.com	Google > Alphabet
8	adform.net	AdForm

Table 13: Top domains with lineage and news page occurrence of case 3.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.  
The approved original version of this thesis is available in print at TU Wien Bibliothek.

# Bibliography

- [1] UK Parliament. Disinformation and ‘fake news’: Interim Report - Digital, Culture, Media and Sport Committee - House of Commons. <https://publications.parliament.uk/pa/cm201719/cmselect/cmcmucmeds/363/36303.htm>, 2018. [Online, accessed on 12-November-2019].
- [2] J. Estrada-Jiménez, J. Parra-Arnau, A. Rodríguez-Hoyos, and J. Forné. Online advertising: Analysis of privacy threats and protection approaches. *Computer Communications* 100, page 32–51, 2017.
- [3] D. Searls. Separating advertising’s wheat and chaff. <http://blogs.harvard.edu/doc/2015/08/12/separating-advertisings-wheat-and-chaff/>, 2015. [Online, accessed on 15-December-2019].
- [4] A. Ghosh, M. Mahdian, R.P. McAfee, and S. Vassilvitskii. To Match or Not to Match: Economics of Cookie Matching in Online Advertising. *ACM Trans. Econ. Comput.*, 3, 2015.
- [5] P. Papadopoulos, N. Kourtellis, and E. Markatos. Cookie Synchronization: Everything You Always Wanted to Know But Were Afraid to Ask. In *The World Wide Web Conference, WWW ’19*, page 1432–1442, New York, NY, USA, 2019. Association for Computing Machinery.
- [6] M.A. Bashir, S. Arshad, W. Robertson, and C. Wilson. Tracing Information Flows Between Ad Exchanges Using Retargeted Ads. In *25th USENIX Security Symposium (USENIX Security 16)*, pages 481–496, Austin, TX, 2016.
- [7] M. Wlosik. What Is a Data Broker and How Does It Work? <https://clearcode.cc/blog/what-is-data-broker/>, 2019. [Online, accessed on 15-December-2019].
- [8] O. Barbu. Advertising, Microtargeting and Social Media. *Procedia - Social and Behavioral Sciences. International Conference on Communication and Education in Knowledge Society 163*, page 44–49, 2014.
- [9] A.M. Hormozi. Cookies and Privacy. In *Information Systems Security*, volume 13, pages 51–59, 2005.

- [10] G. Acar, C. Eubank, S. Englehardt, M. Juarez, A. Narayanan, and C. Diaz. The Web Never Forgets: Persistent Tracking Mechanisms in the Wild. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, CCS '14*, page 674–689, New York, NY, USA, 2014. Association for Computing Machinery.
- [11] N. Kaur, S. Azam, K. Kannoorpatti, K.C. Yeo, and B. Shanmugam. Browser Fingerprinting as user tracking technology. In *2017 11th International Conference on Intelligent Systems and Control (ISCO)*, pages 103–111, 2017.
- [12] R. Upathilake, Y. Li, and A. Matrawy. A classification of web browser fingerprinting techniques. In *2015 7th International Conference on New Technologies, Mobility and Security (NTMS)*, page 1–5, 2015.
- [13] M. Ayenson, D. Wambach, A. Soltani, N. Good, and C. Hoofnagle. Flash cookies and privacy II: now with HTML5 and ETag respawning. *SSRN Electronic Journal*, 2011.
- [14] P. Laperdrix, N. Bielova, B. Baudry, and G. Avoine. Browser Fingerprinting: A survey. *arXiv:1905.01051 [cs]*, 2019.
- [15] P. Eckersley. How Unique Is Your Web Browser? In *Privacy Enhancing Technologies*, page 1–18, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.
- [16] J. Ruohonen and V. Leppänen. Invisible Pixels Are Dead, Long Live Invisible Pixels! In *Proceedings of the 2018 Workshop on Privacy in the Electronic Society, WPES'18*, page 28–32, New York, NY, USA, 2018. Association for Computing Machinery.
- [17] C. Wang, B. Zhu, and M. Zuo. INTEGRTING DIFFERENT TYPES OF TARGETING METHODS IN ONLINE ADVERTISING. 2016. PACIS 2016 Proceedings. 303.
- [18] M. Zawadziński. AdTech Targeting Methods: The Ultimate Guide. <https://clearcode.cc/blog/adtech-targeting-guide/>, 2016. [Online, accessed on 15-January-2020].
- [19] F. Zuiderveen Borgesius, J. Moeller, S. Kruikemeier, R. Ó Fathaigh, K. Irion, T. Dobber, B. Bodó, and C.H. de Vreese. Online Political Microtargeting: Promises and Threats for Democracy. *Utrecht Law Review*, 14(1), page 82–96, 2018.
- [20] Y. Kim, J. Hsu, D. Neiman, C. Kou, L. Bankston, S. Kim, R. Heinrich, R. Baragwanath, and G. Raskutti. The Stealth Media? Groups and Targets behind Divisive Issue Campaigns on Facebook. *Political Communication* 35, page 1–29, 2018.
- [21] UK Parliament. Ads supplied by Facebook to the DCMS Committee. [https://www.parliament.uk/documents/commons-committees/culture-media-and-sport/Fake\\_news\\_evidence/Ads-supplied-by-](https://www.parliament.uk/documents/commons-committees/culture-media-and-sport/Fake_news_evidence/Ads-supplied-by-)

Facebook-to-the-DCMS-Committee.pdf. [Online, accessed on 12-November-2019].

- [22] UK Parliament. BrexitCentral/BeLeave Ads. [https://www.parliament.uk/documents/commons-committees/culture-media-and-sport/Fake\\_news\\_evidence/Brexit-Central-BeLeave-Ads.pdf](https://www.parliament.uk/documents/commons-committees/culture-media-and-sport/Fake_news_evidence/Brexit-Central-BeLeave-Ads.pdf). [Online, accessed on 12-November-2019].
- [23] UK Parliament. Vote Leave/50 Million Ads. [https://www.parliament.uk/documents/commons-committees/culture-media-and-sport/Fake\\_news\\_evidence/Vote-Leave-50-Million-Ads.pdf](https://www.parliament.uk/documents/commons-committees/culture-media-and-sport/Fake_news_evidence/Vote-Leave-50-Million-Ads.pdf). [Online, accessed on 12-November-2019].
- [24] The Guardian. The Guardian view on the Leave campaign: show some respect for truth | Editorial. <https://www.theguardian.com/commentisfree/2016/may/27/the-guardian-view-on-the-leave-campaign-show-some-respect-for-truth>, 2016. [Online, accessed on 06-January-2020].
- [25] UK Parliament. Brexit Central BeLeave Spreadsheet. [https://www.parliament.uk/documents/commons-committees/culture-media-and-sport/Fake\\_news\\_evidence/Brexit-Central-BeLeave-Spreadsheet.pdf](https://www.parliament.uk/documents/commons-committees/culture-media-and-sport/Fake_news_evidence/Brexit-Central-BeLeave-Spreadsheet.pdf). [Online, accessed on 12-November-2019].
- [26] UK Parliament. Vote Leave 50 million spreadsheet. [https://www.parliament.uk/documents/commons-committees/culture-media-and-sport/Fake\\_news\\_evidence/Vote-Leave-50-million-spreadsheet.pdf](https://www.parliament.uk/documents/commons-committees/culture-media-and-sport/Fake_news_evidence/Vote-Leave-50-million-spreadsheet.pdf). [Online, accessed on 12-November-2019].
- [27] J. Stone. British public still believe Vote Leave ‘£350m a week to EU’ myth from Brexit referendum. <https://www.independent.co.uk/news/uk/politics/vote-leave-brexit-lies-eu-pay-money-remain-poll-boris-johnson-a8603646.html>, 2018. [Online, accessed on 06-January-2020].
- [28] OECD. Youth unemployment rate (indicator). <https://data.oecd.org/unemp/youth-unemployment-rate.htm>, 2019. [Online, accessed on 12-November-2019].
- [29] Turkey Threatens to Suspend EU Migrant Deal Due to Lack of Visa-free Travel to EU. <https://www.schengenvisa.info.com/news/turkey-threatens-to-suspend-eu-migrant-deal-due-to-lack-of-visa-free-travel-to-eu/>, 2019. [Online, accessed on 16-January-2020].
- [30] M. Holehouse. EU to launch kettle and toaster crackdown after Brexit vote. <https://www.telegraph.co.uk/news/2016/05/10/eu-to-launch-kettle-and-toaster-crackdown-after-brexit-vote2/>, 2016. [Online, accessed on 16-January-2020].

- [31] UK Parliament. Russian influence in political campaigns. <https://publications.parliament.uk/pa/cm201719/cmselect/cmcomeds/363/36308.htm>, 2018. [Online, accessed on 06-January-2020].
- [32] G. Hinsliff. If Russia meddled in the Brexit vote we need to know – before the election. <https://www.theguardian.com/commentisfree/2019/nov/05/russia-brexit-vote-election-boris-johnson-intelligence-committee-report>, 2019. [Online, accessed on 06-January-2020].
- [33] C. Cadwalladr and E. Graham-Harrison. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>, 2018. [Online, accessed on 15-February-2020].
- [34] J. Gottfried and E. Shearer. News Use Across Social Media Platforms 2016. <https://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/>, 2016. [Online, accessed on 06-January-2020].
- [35] O. Solon. Facebook’s failure: did fake news and polarized politics get Trump elected? <https://www.theguardian.com/technology/2016/nov/10/facebook-fake-news-election-conspiracy-theories>, 2016. [Online, accessed on 06-January-2020].
- [36] C. Shao, G.L. Ciampaglia, A. Flammini, and F. Menczer. Hoaxy: A Platform for Tracking Online Misinformation. *Proceedings of the 25th International Conference Companion on World Wide Web*, page 745–750, 2016.
- [37] E. Helmore. Facebook commitment to free speech will ‘piss people off’, Zuckerberg says. <https://www.theguardian.com/technology/2020/feb/01/facebook-political-ads-zuckerberg>, 2020. [Online, accessed on 15-February-2020].
- [38] European Commission. European Commission calls on national political parties to join efforts to ensure free and fair elections in Europe. [https://ec.europa.eu/commission/presscorner/detail/en/IP\\_19\\_1672](https://ec.europa.eu/commission/presscorner/detail/en/IP_19_1672), 2018. [Online, accessed on 16-February-2020].
- [39] Facebook. Protecting Elections in the EU. <https://about.fb.com/news/2019/03/ads-transparency-in-the-eu/>, 2019. [Online, accessed on 16-February-2020].
- [40] Facebook. Werbebericht. <https://www.facebook.com/ads/archive/report/>. [Online, accessed on 16-February-2020].



- [41] Facebook. Facebook advertising targeting options. <https://www.facebook.com/business/ads/ad-targeting>. [Online, accessed on 16-January-2020].
- [42] General Data Protection Regulation (GDPR) - Consent. <https://gdpr-info.eu/issues/consent/>. [Online, accessed on 16-February-2020].
- [43] Cookies, the GDPR, and the ePrivacy Directive. <https://gdpr.eu/cookies/>, 2019. [Online, accessed on 16-January-2020].
- [44] Recital 30 - Online identifiers for profiling and identification. <https://gdpr.eu/recital-30-online-identifiers-for-profiling-and-identification/>, 2018. [Online, accessed on 16-January-2020].
- [45] X. Hu and N. Sastry. Characterising Third Party Cookie Usage in the EU after GDPR. In *Proceedings of the 10th ACM Conference on Web Science, WebSci '19*, page 137–141, New York, NY, USA, 2019. Association for Computing Machinery.
- [46] A. K. Tantleff, S. Millendorf, and R. E. Glass. Top European Court Rules Pre-Checked Cookie Consent Boxes Invalid. <https://www.natlawreview.com/article/top-european-court-rules-pre-checked-cookie-consent-boxes-invalid>, 2019. [Online, accessed on 6-December-2019].
- [47] Privacy Policy | ResearchGate, the professional network for scientists. <https://www.researchgate.net/privacy-policy>, 2019. [Online, accessed on 07-November-2019].
- [48] C.M. Gray, Y. Kou, B. Battles, J. Hoggatt, and A.L. Toombs. The Dark (Patterns) Side of UX Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18*, New York, NY, USA, 2018. Association for Computing Machinery.
- [49] H. Brignull and A. Darlo. Types of Dark Pattern. <https://www.darkpatterns.org/types-of-dark-pattern>, 2019. [Online, accessed on 14-December-2019].
- [50] T. Vieira. Dark patterns: Welcome to the dark side of a user-friendly industry. <https://blog.prototypr.io/dark-patterns-welcome-to-the-dark-side-of-a-user-friendly-industry-ef78ac95a5be>, 2019. [Online, accessed on 14-December-2019].
- [51] C. Utz, M. Degeling, S. Fahl, F. Schaub, and T. Holz. (Un)Informed Consent: Studying GDPR Consent Notices in the Field. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, CCS '19*, page 973–990, New York, NY, USA, 2019. Association for Computing Machinery.
- [52] A. Fowler. Lightbeam for Firefox: Privacy Education for Users & Open Data for Publishers. <https://blog.mozilla.org/blog/2013/10/25/>

lightbeam-for-firefox-privacy-education-for-users-open-data-for-publishers/, 2013. [Online, accessed on 01-November-2019].

- [53] MyShadow. Trackography | Me and my Shadow. <https://myshadow.org/trackography#what-is-trackography>, 2017. [Online, accessed on 07-November-2019].
- [54] Tor Project | Anonymity Online. <https://www.torproject.org/>. [Online, accessed on 15-November-2019].
- [55] Privacy Badger. PRIVACY BADGER FAQ | Electronic Frontier Foundation. <https://www.eff.org/privacybadger/faq>. [Online, accessed on 10-November-2019].
- [56] A. Esteve. The business of personal data: Google, Facebook, and privacy issues in the EU and the USA. *International Data Privacy Law* 7, page 36–47, 2017.
- [57] Facebook, Inc. - ANNUAL REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934. <http://d18rn0p25nwr6d.cloudfront.net/CIK-0001326801/45290cc0-656d-4a88-a2f3-147c8de86506.pdf>, 2019. [Online, accessed on 05-February-2020].
- [58] Alphabet Inc. - ANNUAL REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934. [https://abc.xyz/investor/static/pdf/20200204\\_alphabet\\_10K.pdf](https://abc.xyz/investor/static/pdf/20200204_alphabet_10K.pdf), 2019. [Online, accessed on 05-February-2020].
- [59] A.-R. Jung. The influence of perceived ad relevance on social media advertising: An empirical examination of a mediating role of privacy concern. 70:303–309, 2017.
- [60] J. Davies. After GDPR, The New York Times cut off ad exchanges in Europe - and kept growing ad revenue. <https://digiday.com/media/gumgumtest-new-york-times-gdpr-cut-off-ad-exchanges-europe-ad-revenue/>, 2019. [Online, accessed on 16-January-2020].
- [61] MindTake. Ranking der Top 10 beliebtesten Zeitungs-Websites in Österreich im September 2019. <https://de.statista.com/statistik/daten/studie/471326/umfrage/beliebteste-online-zeitungen-in-oesterreich/>, 2019. Statista GmbH. [Online, accessed on 17-November-2019].
- [62] T. Libert. webxray. <https://github.com/timlib/webxray>. [Online, accessed on 17-November-2019].
- [63] T. Libert and R. Binns. Good News for People Who Love Bad News: Centralization, Privacy, and Transparency on US News Sites. In *Proceedings of the 10th ACM Conference on Web Science, WebSci '19*, page 155–164, New York, NY, USA, 2019. Association for Computing Machinery.