

Legal issues of user tracking technologies

An analysis of web and mobile tracking methods utilised by Austrian enterprises for business purposes

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieurin

im Rahmen des Studiums

Business Informatics

eingereicht von

Tanja Travnicek, BSc
Matrikelnummer 01126286

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: ao.Univ.-Prof. Mag. Dr.iur. Markus Haslinger

Wien, 16. April 2022

Tanja Travnicek

Markus Haslinger

Legal issues of user tracking technologies

An analysis of web and mobile tracking methods utilised by Austrian enterprises for business purposes

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieurin

in

Business Informatics

by

Tanja Travnicek, BSc

Registration Number 01126286

to the Faculty of Informatics

at the TU Wien

Advisor: ao.Univ.-Prof. Mag. Dr.iur. Markus Haslinger

Vienna, 16th April, 2022

Tanja Travnicek

Markus Haslinger

Erklärung zur Verfassung der Arbeit

Tanja Travnicek, BSc

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, 16. April 2022

Tanja Travnicek

Acknowledgements

I wish to thank various people for their contribution to this project; Mr Matthias Fassl for his suggestions on relevant research; Mr Michael Langowski for proofreading; and the editorial staff of ".trend" magazine for their willingness to share their data on Austria's Top 500 enterprises.

Special thanks should be given to ao.Univ.-Prof. Mag. Dr.iur. Markus Haslinger, my supervisor, for his professional guidance and constructive input during the planning and development of this work.

I also wish to thank those members of the research community providing access to their work for free in their continuing bid to advance the availability of knowledge to their fellow researchers.

An honorary mention should be given to my cat Pixel for constantly providing distractions and keeping me from finishing this thesis on time.

Finally, I wish to thank my parents for their continued support and encouragement throughout my study.



Diese Arbeit wurde mit einem netidee Stipendium gefördert.

Kurzfassung

Ziel dieser Arbeit ist es, ein klareres Bild der bestehenden Forschung zur Nutzung und zu rechtlichen Aspekten von User-Tracking-Methoden zu gewinnen.

Informationen zum Stand der Technik im Bereich Web- und Mobile-Tracking werden aus einer Reihe fragmentierter Forschungsbereiche zusammengefasst. Diese Arbeit zeigt deutlich, dass österreichische Unternehmen eine Vielzahl an User-Tracking-Methoden für geschäftliche Zwecke einsetzen. Die durchgeführte Umfrage ebenso wie die Ergebnisse des Web-Scraping-Prozesses zeigen, unter anderem, den Einsatz von Erst- und Drittanbieter-Cookies, anderen DOM-Speichermethoden, Tracking über JS-Dateien und Fingerprinting-Verfahren. Die Arbeit weist auch nach, dass ein erheblicher Teil des User-Trackings durch Tools von Drittanbietern erfolgt, insbesondere von großen U.S.-Online-Plattformen wie Google und Facebook. Die durchgeführte rechtliche Analyse offenbart, dass viele der fraglichen Tracking-Methoden bereits durch österreichisches bzw. EU-Recht geregelt sind. Allerdings stellt diese Arbeit auch fest, dass viele Unternehmen noch nicht alle Regelungen volumnfänglich anwenden.

Diese Arbeit liefert ebenso einen klaren Beweis dafür, dass die österreichische Umsetzung von Art. 5 Abs. 3 der Datenschutzrichtlinie für elektronische Kommunikation die Anwendbarkeit derselben auf bestimmte Arten des User-Trackings schmälert. Außerdem zeigt sie, dass österreichische Unternehmen folglich an ein laxeres Regelwerk gebunden sind als andere europäische Unternehmen. Darüber hinaus enthält sie Beispiele, anhand derer gezeigt werden kann, dass Drittanbieter den jeweiligen Erstanbieter teilweise falsche Angaben machen und eine sachgerechte Risikobewertung verunmöglichen. Dies erschwert die legale Nutzung der entsprechenden Tools erheblich.

Abstract

The aim of this thesis is to provide a clearer picture on existing research on utilisation and legal aspects of user tracking methods.

Information is gathered from a number of fragmented research areas on the state-of-the-art in web and mobile tracking. This thesis clearly shows that Austrian enterprises employ a wide range of user tracking methods for business purposes. The conducted survey as well as the employed web scraping reveals, among others, the utilisation of first- and third-party cookies, other DOM storage methods, tracking via JS files and fingerprinting methods. It also shows that a significant number of user tracking is done through tools provided by third parties, especially major U.S. online platforms, like Google and Facebook. The performed legal analysis reveals that many of the tracking methods in question are already regulated by Austrian or EU law. However, this thesis also finds that many companies do not yet apply all of the regulations to their full extent.

This thesis also provides clear evidence that the Austrian transposition of Article 5(3) of the Directive on privacy and electronic communications reduces its applicability to several user tracking methods and that therefore Austrian enterprises are held to a laxer set of rules than other European companies. In addition, it shows that third parties in some cases give false information to the respective first parties and do not allow for a proper risk assessment. This hinders the lawful utilisation of such tools significantly.

Contents

Kurzfassung	ix
Abstract	xi
Contents	xiii
1 Introduction	1
2 Literature review: User tracking	3
2.1 Methodology	3
2.2 History & Definition	10
2.3 Methods & Mechanisms	12
3 Survey: Utilisation of user tracking technologies by Top 500 Austrian Enterprises	33
3.1 Methodology & Questionnaire design	33
3.2 Implementation & Distribution	41
3.3 Results	43
4 Web scraping: Utilisation of user tracking technologies by Top 500 Austrian Enterprises	47
4.1 Methodology & Definition	47
4.2 Implementation & Execution	49
4.3 Results	52
5 Legal analysis: User tracking in Austria and the European Union	67
5.1 Methodology	67
5.2 Overview of relevant legal texts	68
5.3 Applicability & Interpretation	74
6 Summary and future work	103
6.1 Summary	103
6.2 Future work	104

A Appendix: Survey	107
A.1 Lists: Common user tracking tools	107
A.2 Online questionnaire	110
List of Figures	127
List of Listings	129
List of Legal texts	131
Acronyms	133
Bibliography	135

1

CHAPTER

Introduction

A persons right to privacy as well as their right to protection of personal data are granted by the Charter of Fundamental Rights of the European Union.¹ The European Union (EU) has issued a number of legal texts trying to ensure those rights. The most prominent of the aforementioned texts is probably the General Data Protection Regulation (GDPR). Its coming into effect in 2018 has caused a stir. United States (of America) (U.S.) companies were suddenly blocking traffic coming from the EU² and people who have given little thought to data protection before started talking about it.³ Despite the sudden rise in interest about data protection, many people are still unaware that personal data has become a major source of economic value.⁴ In 2018, Malgieri and Custers proposed explicitly informing users about the monetary value of their personal data whenever such data is collected to raise awareness.⁵ Monetary value is mostly established indirectly, e.g. by how useful the data in question is to companies which are trying to target potential customers with personalised advertisements.⁶

By establishing economic value of personal data, it was assumed that Austrian companies collecting personal data for business purposes exist. This thesis aims to validate said assumption by performing a technical survey of user tracking methods utilised by Austrian enterprises, especially focusing on web and mobile tracking methods. Lastly, coming back to the legal aspect of personal data, methods used were analysed considering their legal implications. To the best of my knowledge, there is no other scientific work focusing on these matters with an emphasis on Austria and Austrian enterprises up until the present time.

¹[Charter of Fundamental Rights of the European Union, 2012], Art. 7 & 8

²<https://www.bbc.com/news/world-europe-44248448>, last visited 2021-12-05

³<https://www.irishtimes.com/business/everyone-s-talking-about-gdpr-but-who-s-got-their-head-around-it-1.3474346>, last visited 2021-12-05

⁴[Esteve, 2017], p. 36

⁵[Malgieri and Custers, 2018], p. 290

⁶[Malgieri and Custers, 2018], p. 296

1. INTRODUCTION

The amount of methods and technologies used for data collection as well as the fragmented academic literature on those topics⁷ made it necessary to formulate the following question:

1st research question:

What is currently considered state-of-the-art user tracking in regard to web or mobile phones, and what other methods are (still) in use?

It is answered in chapter 2. Section 2.1 provides insight into the utilised research method, section 2.2 establishes a definition as well as a short history of user tracking and in the last section of chapter 2 an overview of uncovered user tracking methods is given.

Based on the information contained in chapter 2, an online survey was designed. It was intended to answer the following question:

2nd research question:

Which (state-of-the-art) user tracking technologies are currently being used by Austrian enterprises for business purposes?

An overview on the survey design process, along with a report of the survey's implementation and results thereof can be found in chapter 3.

In order to balance out potential shortcomings of the results of chapter 3, another approach was taken in chapter 4 to gather conclusive data about Austrian enterprises and their use of web tracking methods: Web scraping. General information about the approach and its limitations can be found in section 4.1, used tools and the actual implementation are outlined in section 4.2 and the gathered results are discussed in section 4.3.

Finally, the legal aspect of this thesis is subsumed in the following question:

3rd research question:

In which ways are these technologies covered by Austrian or European law? Which technologies in use pose potential threats to users' privacy and are currently not regulated?

With the aim of answering this question all information gathered in the preceding chapters is combined with relevant legislation and court rulings in a legal analysis in chapter 5. The methods used to establish which legal texts are applicable or otherwise relevant, are discussed in section 5.1 and a short overview about said texts is given in section 5.2. The concrete legal implications of the methods found in chapter 3 and chapter 4 are discussed in section 5.3.

The thesis concludes with a summary of findings and an outlook on possible future work in chapter 6.

⁷[Christl et al., 2017], p. 6 f.

CHAPTER 2

Literature review: User tracking

2.1 Methodology

To evaluate the state-of-the-art, as posed in the first rather broad research question, a semi-systematic literature review was applied. This method was suggested by Hanna Snyder in her paper on literature review¹ for exactly this kind of questions. Snyder puts forward a paper by Wong et al.² as a basis for a semi-systematic literature review. The groundwork for Wong et al.'s paper is the methodology of meta-narrative review developed by Greenhalgh et al. in 2005.³ Aforementioned methodology is designed for whole teams of researchers and therefore is not entirely suitable for a master thesis done by a single person. For this reason, some of the proposed review phases⁴ were reduced or left out entirely. Items of the following list, which are written in standard font, were done to their full extent; items written in cursive font were done in reduced or slightly altered form and crossed out parts were left out entirely. The list itself is copied from Greenhalgh et al.'s 2005 paper and shortened for readability.

1. Planning phase

- (a) ~~Assemble multidisciplinary team~~
- (b) Outline initial research question in a broad, open-ended format
- (c) ~~Agree outputs with funder/client~~
- (d) *Face-to-face review meetings, including planned input from external peers*

¹[Snyder, 2019], p. 334

²[Wong et al., 2013]

³[Greenhalgh et al., 2005]

⁴[Greenhalgh et al., 2005], p. 420

2. LITERATURE REVIEW: USER TRACKING

2. Search phase

- (a) Initial search led by intuition, informal networking and "browsing"
- (b) Search for seminal conceptual papers by tracking references of references. Evaluate those references by scholarship, comprehensiveness and contribution to subsequent work.
- (c) *Search for empirical papers by electronic searching key databases, hand searching key journals and "snowballing"*

3. Mapping phase

Identify:

- (a) *Key elements of the research paradigm (conceptual, theoretical, methodological and instrumental)*
- (b) *Key actors and events*
- (c) ~~Prevailing language and imagery used by scientists to "tell the story"~~

4. Appraisal phase

Using appropriate critical appraisal techniques:

- (a) Evaluate each primary study for its validity and relevance to the review question
- (b) Extract and collate key results, grouping comparable studies together

5. Synthesis phase

- (a) Identify all key dimensions of the problem that have been researched
- (b) *For each dimension, give a narrative account of the contribution (if any)*
- (c) *Treat conflicting findings as higher-order data and explain in terms of contestation between the different paradigms from which the data were generated*

6. Recommendation phase

- (a) Summarise the overall messages along with other relevant evidence
- (b) ~~Distil and discuss recommendations for practice, policy and further research~~

Greenhalgh et al. did a study to show which results this methodology will yield. Their most important finding was that, although their review methodology is only part-complete, their peers agreed that their study illuminated and clarified a previously confusing literature.⁵

The following sections provide insight into the application of the aforementioned methodology. For the results skip to section 2.3.

⁵[Greenhalgh et al., 2005], p. 427

2.1.1 Planning phase

As outlined in chapter 1, the initial problem was broken down in three individual research questions. Chapter 2 focuses on the first question only. The question was worded in a broad and open-ended format to fulfill item (b) of the planning phase.

To partially adhere to item (d) of the planning phase, the question has been discussed with my supervisor. It was also presented to the dean of the Faculty of Informatics, TU Wien, the dean of study for Business Informatics at the Faculty of Informatics, TU Wien and fellow Master students in an online-meeting in September 2020. Two of my peers had been tasked with reviewing my thesis proposal which preceded this work. They concluded the selected research question to be suitable.

2.1.2 Search phase

Item (a) of the search phase asks for intuition. Therefore, I started researching a definition and short history of user tracking methods and technologies. Knowledge of a topic's past often helps to determine what should be considered state-of-the-art.

My supervisor pointed out the work of Wolfie Christl, Director of Cracked Labs - Institute for Critical Digital Culture⁶. "*Cracked Labs is an independent research institute and a creative laboratory based in Vienna, Austria. It investigates the socio-cultural impacts of information technology and develops social innovations in the field of digital culture.*"⁷ Christl published several books on user tracking methods used by companies and corporations.

Additional papers on cookies, tracking and consent notices were suggested by Matthias Fassl, a Doctoral Researcher at the CISPA Helmholtz Center for Information Security⁸. Together with the papers used as references for my thesis proposal, they concluded the first step of the search phase.

In the next step (item (b) of the search phase), I referred to the references included in the aforementioned papers. A visual representation of this search phase step can be found in fig. 2.1.

The chapter "Recording Personal Data – Devices and Platforms" of Christl and Spiekermann's book "Networks of Control"⁹ was especially helpful in this regard. While combing through aforementioned chapter and gathering interesting references, two references appeared to be missing ("(see Ackerman 2013)"¹⁰ and "(see Sterbik-Lamina et al 2009)".¹¹ They were provided by Christl after a short email correspondence.

⁶<https://wolfie.crackedlabs.org/en>, last visited 2020-12-29

⁷<https://crackedlabs.org/en>, last visited 2021-01-04

⁸<https://cispa.de/en/people/matthias.fassl>, last visited 2020-12-29

⁹[Christl and Spiekermann, 2016], p. 45-75

¹⁰[Christl and Spiekermann, 2016], p. 64

¹¹[Christl and Spiekermann, 2016], p. 70

2. LITERATURE REVIEW: USER TRACKING

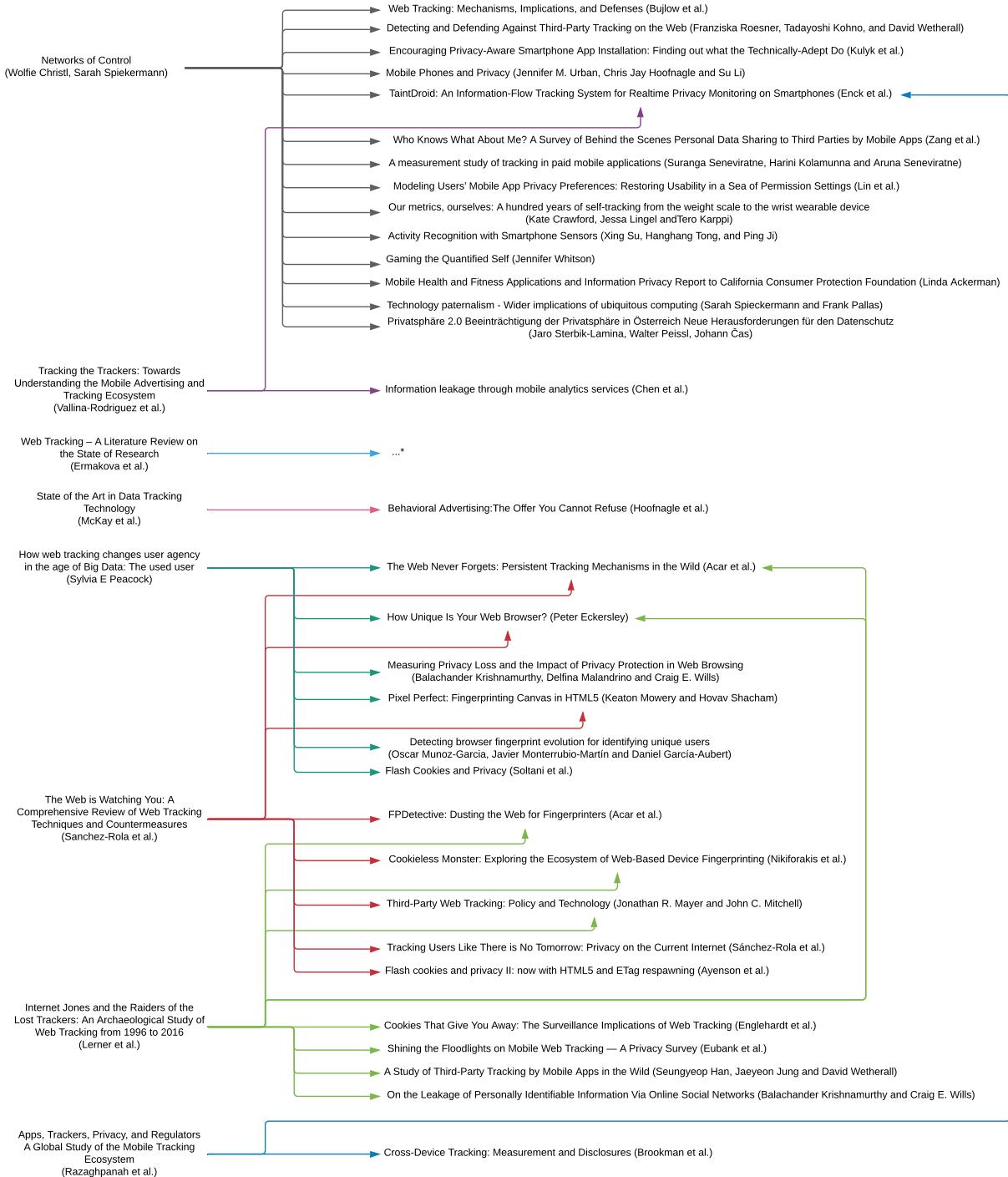


Figure 2.1: Visualisation of the search process for item (b) of the search phase

The second notable publication is "Web Tracking – A Literature Review on the State of Research" by Ermakova et al.¹² Most of their 61 references are related to the first research question of this thesis. Due to time constraints and because Ermakova et al. already did a literature review, a decision to include only Ermakova et al.'s paper and not the referenced original papers into the review pool was made. This is why no search results are listed for this paper in fig. 2.1

While comparing references, Sanchez-Rola et al.'s "The Web is Watching You: A Comprehensive Review of Web Tracking Techniques and Countermeasures"¹³, Peacock's "How web tracking changes user agency in the age of Big Data: The used user"¹⁴ and Lerner et al.'s "Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study of Web Tracking from 1996 to 2016"¹⁵ had the most references in common. This can also be seen in fig. 2.1.

The papers cited the most by publications in the initial review pool were Acar et al.'s paper on "The Web Never Forgets: Persistent Tracking Mechanisms in the Wild"¹⁶ together with Eckersley's "How Unique Is Your Web Browser?"¹⁷ and Enck et al.'s "TaintDroid: an information-flow tracking system for realtime privacy monitoring on smartphones".¹⁸

Due to the success of the first two steps within the search phase, the third step was reduced to a single query on Google Scholar.¹⁹ A search for *Literature review "user tracking"* resulted in 992 hits published 2018 or later, excluding patents and citations. Of those 55 were deemed relevant to this thesis based on their title and kept for further investigation.

2.1.3 Mapping phase & Appraisal phase

Although Greenhalgh et al. keep mapping and appraisal phases separate,²⁰ those phases were merged for this work. "*Validity and relevance to the research question*" (Appraisal phase, item (a)) was determined while looking for information on "*key actors*", "*key events*" (both part of Mapping phase, item (b)) and "*key elements of the research paradigm*" (Mapping phase, item (a)). The key findings of the mapping phase were used to group work together (Appraisal phase, item (b)). To aid this part of the process, papers gathered in the search phase were now tagged with their year of publication, their authors, their authors' affiliation and one or more keywords regarding their content. Tags were created and managed using Evernote.²¹ Each paper was represented as a note in an Evernote notebook and tagged accordingly.

¹²[Ermakova et al., 2018]

¹³[Sanchez-Rola et al., 2016]

¹⁴[Peacock, 2014]

¹⁵[Lerner et al., 2016]

¹⁶[Acar et al., 2014]

¹⁷[Eckersley, 2010]

¹⁸[Enck et al., 2010]

¹⁹<https://scholar.google.com/>, last visited 2021-01-13

²⁰[Greenhalgh et al., 2005], p. 420

²¹<https://evernote.com/>, last visited 2021-01-13

2. LITERATURE REVIEW: USER TRACKING

Publications deemed non-relevant in the tagging process were excluded immediately. This was the case for 17 publications from the Google Scholar results. That number includes publications which were excluded due to their scientifically questionable status, e.g. papers published without peer-review. Furthermore, twelve publications were considered potentially non-relevant, but were tagged nonetheless. Of these, five were later excluded due to their publishing date before 2010.

Seven publications found while querying Google Scholar were considered non-relevant to this chapter but kept because they contain relevant information for later chapters. For these no notes or tags were created.

After the tagging process, a total of 63 publications were left. The following figures and statistical analyses are included for better understanding and each visualise or explains a certain metric of the gathered publications.

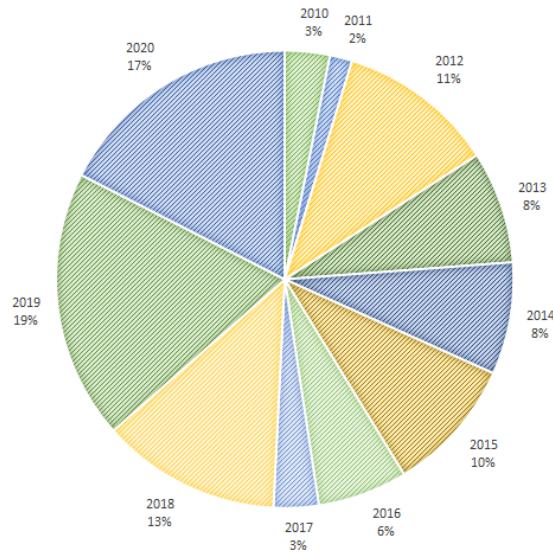


Figure 2.2: Tagged publications grouped by publication year

The authors with the most contributions were Arvind Narayanan (4 publications), Iskander Sanchez-Rola (4 publications), Igor Santos (4 publications), Gunes Acar (3 publications), Christian Eubank (3 publications), Chris Jay Hoofnagle (3 publications) and Narseo Vallina-Rodriguez (3 publications).

Size and colour of the bubbles in fig. 2.3 indicate the total number of publications featuring a certain tag. Their proximity shows how often tags were used together. Research on privacy enhancing technologies often tries to prevent very particular ways of tracking, which are not described in so much detail elsewhere. Therefore, these publications are included in this chapter although their focus is in clear contrast to most other publications on user tracking.

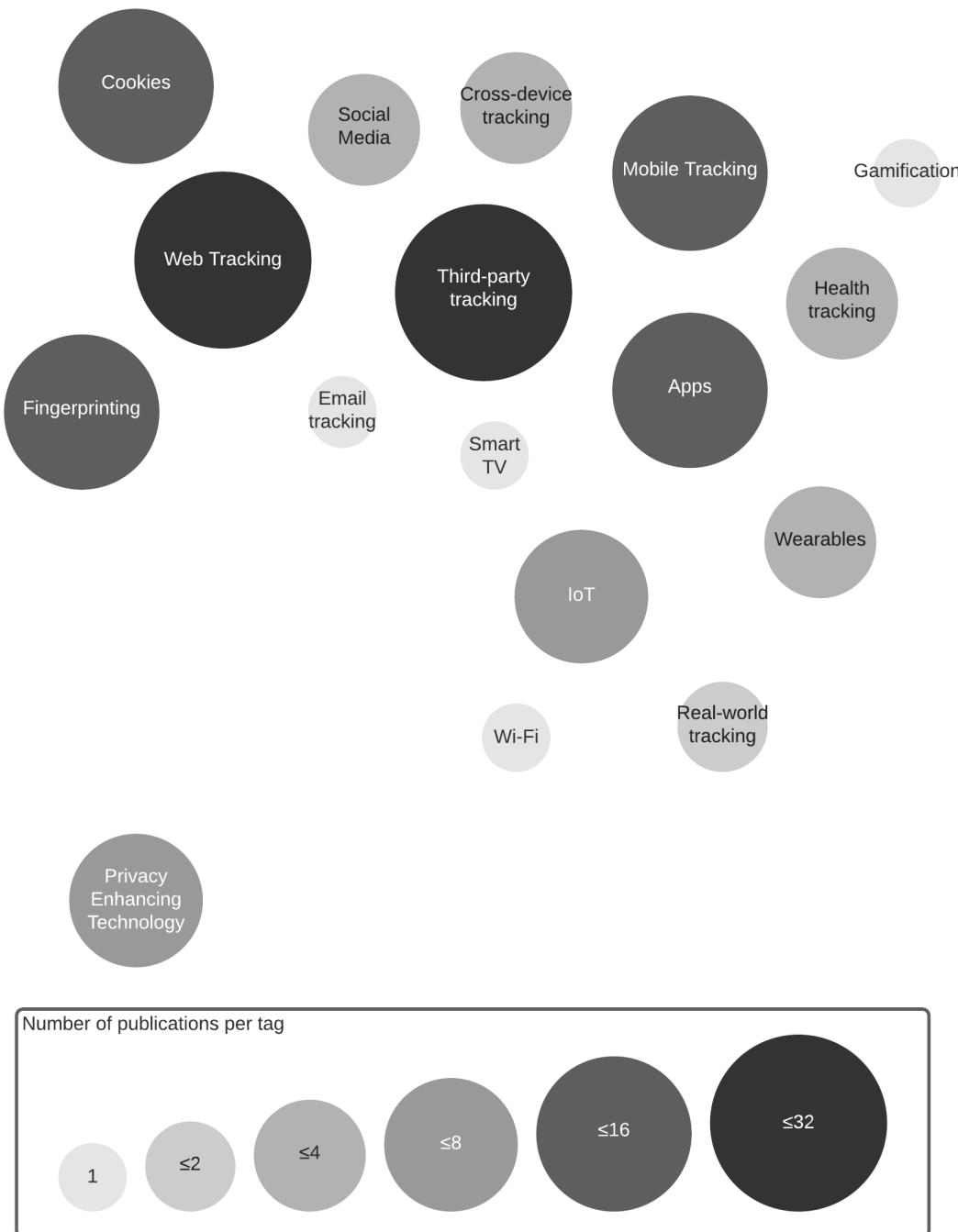


Figure 2.3: Tagged publications grouped by content

2.1.4 Synthesis & Recommendation phase

The results of synthesis and recommendation phase can be found in the following sections of this chapter. In this process a number of papers were added either when a section was lacking in important information or when new aspects of a certain topic became available, e.g. the paper on DNS-based Tracking Evasion.²²

2.2 History & Definition

2.2.1 Definition

There is no clear definition of user tracking as there are many ways how, when and why to track users. When one studies available information about user tracking, one even gets the impression, that there is more data collected about users than anyone is currently able to purposefully process.

In addition to missing a clear definition, there are different terms with similar meaning in use, like "consumer tracking" or "customer tracking".²³ McKay et al. use the term "data gathering" in their 2019 report on the "State of the Art in Data Tracking Technology".²⁴ Also, "web tracking" and "user tracking" are often used interchangeably in the anglophone world, although web tracking does only cover part of user tracking practices.²⁵

In general, two types of user tracking are distinguished: first- and third-party tracking. First-party tracking is done by the service provider a user is directly interacting with, e.g. visiting their website or using their app. Third-party tracking is done by additional parties present within the first party's services, e.g. through libraries included in a first-party application, or by parties which are handed information by the first party in other ways.

2.2.2 History

According to McKay et al. user tracking dates back to the late 1970s, when Texas International Airlines started gathering information about their frequent flyers.²⁶ Therefore, user tracking pre-dates the commercial phase of internet history (1984-1989).²⁷ Bujlow et al. produced a timeline of web tracking mechanisms based on their first documented appearance,²⁸ it can be seen in fig. 2.4. A major step in user tracking was made with the invention of Hypertext Transfer Protocol (HTTP) cookies by Lou Montulli.²⁹

²²[Dimova et al., 2021]

²³<http://www.emarketingdictionary.com/WebMarketingDictionary-Customer-Tracking-Definition.html>, last visited 2021-01-26

²⁴[McKay et al., 2019]

²⁵<https://www.atinternet.com/en/glossary/user-tracking-web-tracking/>, last visited 2021-01-26

²⁶[McKay et al., 2019], p. 4

²⁷[Cohen-Almagor, 2013], p. 19

²⁸[Bujlow et al., 2015], p. 2

²⁹[Montulli, 1995]

Before that, user tracking was limited to single sessions,³⁰ which for example made it impossible to detect re-visiting customers. And even though HTTP cookies have been around since the early 1990s, they still are one of the most used web tracking technologies to date.³¹

They are also the most prominent user tracking method as their use has to be declared on all websites accessible to EU-citizens since May 2011.³² On top of first-party tracking, which is limited to one specific domain, third-party web tracking can be implemented on many websites and *"is typically done for the purposes of website analytics, targeted advertising, and other forms of personalization (e.g., social media content)."³³* Lerner et al. showed how the number of third-party web trackers increased from 1996 to 2016 (see fig. 2.5). Their characterisation of tracking behaviour is more thoroughly explained in section 2.3.5.

At least third-party cookies could soon be a thing of the past. The European Court of Justice clarified some former "grey areas" on consent notices³⁴ and major browsers like Firefox and Safari are already blocking third-party cookies by default. Even Google is moving into the same direction.³⁵

The following section contains information on other tracking methods, which allow for far more detailed user profiles than cookies ever could. They are also *"far less controllable and far more privacy-damaging"*, at least that is how Al-Fannah and Mitchel

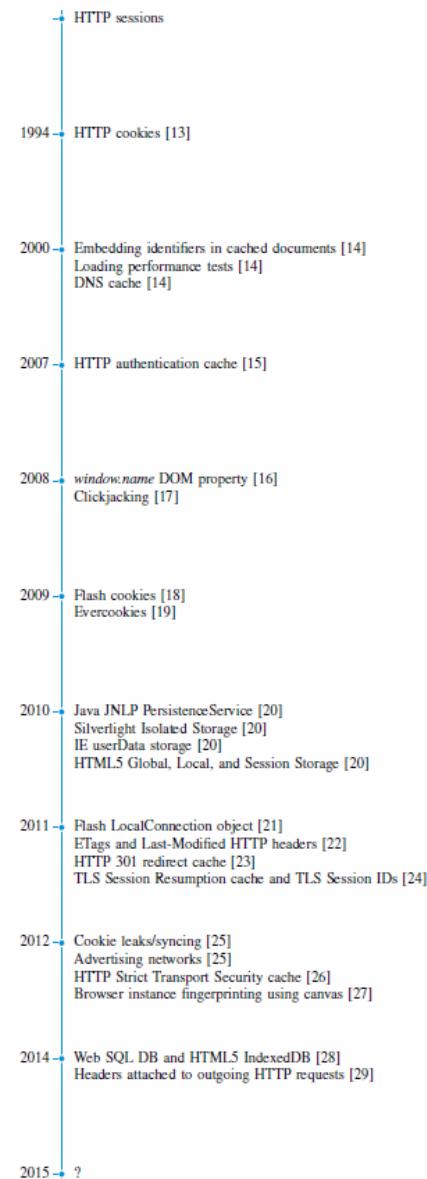


Figure 2.4: History of web tracking by Bujlow et al. from 1990 to 2015²⁸

³⁰[Bujlow et al., 2015], p. 5

³¹[Mallet and Beauvisage, 2019], p. 2

³²[Directive 2009/136/EC, 2009]

³³[Lerner et al., 2016], p.998

³⁴[Planet49 (C-673/17), 2019]

³⁵<https://www.theverge.com/2020/1/14/21064698/google-third-party-cookies-chrome-two-years-privacy-safari-firefox>, last visited 2021-02-09

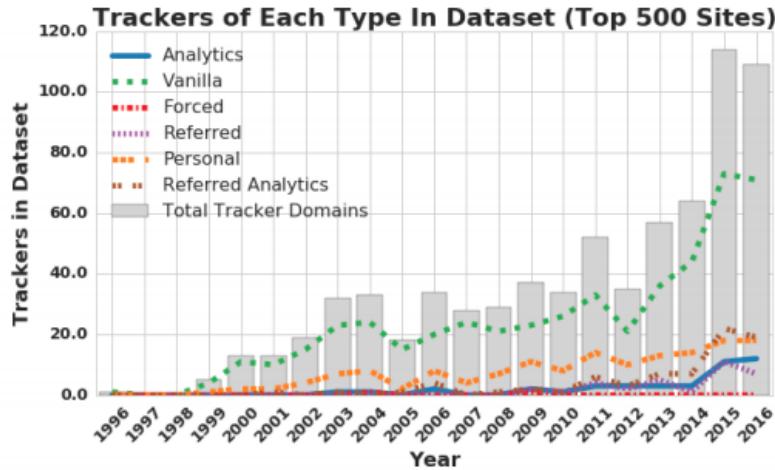


Figure 2.5: Rise in tracking domains from 1996 to 2016 (Domains can have more than one tracking behaviour. Therefore, the sum of lines might be bigger than the bar's height) ³⁷

put it in their 2020 paper on browser fingerprinting.³⁶ Additionally, the current push towards shared identity solutions, see section 2.3.5, could lead to a centralisation of more and more information about a user's behaviour at very few third-party domains.

When Apple sold more than 1 million iPhones in the first five days after its initial release in 2007,³⁸ it opened a whole new market to advanced user tracking methods. Traditional mobile phones were not immune to user tracking, as the Internet could be accessed from mobile phones since the late 1990s,³⁹ but the new additional sensors and functions integrated in smartphones pose an even greater risk to users' privacy.⁴⁰ More details about that can be found in the next section.

2.3 Methods & Mechanisms

2.3.1 Web tracking

In their 2015 paper on "Web Tracking: Mechanisms, Implications, and Defenses" Bujlow et al. introduced five different classifications for web tracking mechanisms. They distinguish between session-only, storage-based, cache-based, fingerprinting and other web tracking mechanisms.⁴¹ This section follows this classification. Mayer and Mitchell distinguish

³⁶[Al-Fannah and Mitchell, 2020], p. 173

³⁷[Lerner et al., 2016], p. 1005

³⁸<https://www.apple.com/newsroom/2007/09/10Apple-Sells-One-Millionth-iPhone/> last visited 2021-12-05

³⁹[Meadows, 2013]

⁴⁰[Ermakova et al., 2018], p. 4737

⁴¹[Bujlow et al., 2015], p. 3

only two types of web tracking, stateful (e.g. cookies) and stateless (e.g. fingerprinting)⁴², depending on where the user-identifying data is stored.

Session-only tracking mechanisms

As already mentioned in section 2.2.2, before the invention of HTTP cookies users could only be tracked by session identifiers (session IDs) passed on with every GET- or POST-request⁴¹. Those session IDs are normally stored as cookies nowadays and not passed on in the URL or as hidden form field anymore. Therefore, session-only tracking will not be discussed any further in this thesis.

Storage-based tracking mechanisms

A list of storage-based tracking mechanisms which are still in use and therefore relevant, even though some are not quite state-of-the-art anymore, can be found below.

A number of relevant publications still mention Flash cookies as part of storage-based tracking mechanisms. Flash cookies, or "local shared objects", were 100KB-sized .sol-files stored on a computer, which allowed for cross-browser tracking, because all instances of the Adobe Flash plugin shared the same storage directory.⁴³ Adobe Flash Player and therefore Adobe Flash plugins have reached their end of life on December 31st 2020.⁴⁴ Therefore, Flash cookies are no longer a tracking option and will not be discussed further in this thesis. The same applies to the Internet explorer userData storage, which was declared obsolete in Internet Explorer 7 but continued to function up to the last version of Internet Explorer.⁴⁵ However, even Microsoft itself fades out support for their former browser across their services,⁴⁶ so it is safe to assume that the Internet Explorer userData storage will no longer be relevant in the future.

HTTP cookies:

HTTP cookies were invented by Lou Montulli and first patented by Netscape Communications Corp. in 1995.⁴⁷ Christl and Spiekermann wrote the following about HTTP cookies in their publication "Networks of Control"⁴⁸, partly citing Bujlow et al.⁴⁹: *"Cookies are "small pieces of data", which are "placed in a browser storage by the web server" (Bujlow et al 2015, p. 5) [Annot.: [Bujlow et al., 2015]]. When a website is visited the first time, a unique identification code can be stored in the cookie file on the user's computer.*

⁴²[Mayer and Mitchell, 2012], p. 421

⁴³[Bujlow et al., 2015], p. 6 & [Ayenson et al., 2011], p. 2

⁴⁴<https://www.adobe.com/products/flashplayer/end-of-life.html>, last visited 2021-02-11

⁴⁵[Bujlow et al., 2015], p. 7

⁴⁶<https://techcommunity.microsoft.com/t5/microsoft-365-blog/microsoft-365-apps-say-farewell-to-internet-explorer-11-and-ba-p/1591666>, last visited 2021-02-11

⁴⁷[Montulli, 1995]

⁴⁸[Christl and Spiekermann, 2016], p. 45

⁴⁹[Bujlow et al., 2015]

2. LITERATURE REVIEW: USER TRACKING

Subsequently, the website can recognize the user across further page visits by accessing this identifier again and again. While session cookies expire when the web browser is closed, persistent cookies can be stored for hours, days or years (see Bujlow et al 2015) [Annot.: [Bujlow et al., 2015]]. Both types can be used for authentication purposes or to remember information entered by the user, such as items in an online shopping cart, but also to track which pages were visited and how a user interacted with the website in the past." Cookies can be easily deleted by users. Over the years there have been several attempts to keep cookies from being deleted⁵⁰ or restore them after deletion.⁵¹ These methods are also known as "evercookie vectors",⁵² because they, much like the original "Evercookie", which was presented by Samy Kamkar in 2010,⁵³ provide options to keep or respawn HTTP cookies on a users' computer. Some of them relied on the Adobe Flash Player and are therefore no longer available. Others, however, are relying on newer technology like HTML5, see next sections.

HTML5 localStorage:

HTML5 localStorage objects can be between 5 to 10 MB⁵⁴ in size, they are stored in a database (e.g. SQLite file) and are persistent by default⁵⁵. This makes the HTML5 localStorage a powerful tracking tool even though it is limited to one browser only.

Bujlow et al. state that "*the localStorage is automatically emptied at the time when the cookies are cleared*"⁵⁶. However, most browsers could communicate this feature more clearly. While cookies are often explicitly mentioned and further information about them is given, HTML localStorage is often subsumed with other stored information under the term "(web)site data", for example see Firefox options.⁵⁷

Ayenson et al. found a few cases in 2011 in which HTML5 localStorage was used to mirror HTTP cookies⁵⁸, and Lerner et al. documented a rise in use of the localStorage application programming interface (API), which can be seen in fig. 2.6⁵⁹. In 2018,

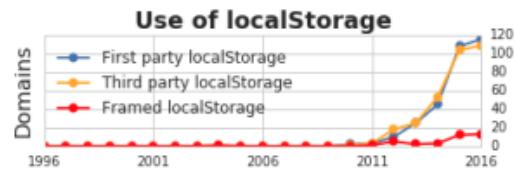


Figure 6: Domains using `window.localStorage`. First party usages are uses in the top frame of a web page by a script loaded from the web page's own domain. Third party usages are those also in the top frame of a page but by a script loaded from a third party. Framed usages are those inside of an iframe.

Figure 2.6: Figure 6 from Lerner et al.'s "Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study of Web Tracking from 1996 to 2016"⁵⁹

⁵⁰[Ayenson et al., 2011], p. 3

⁵¹[Ayenson et al., 2011], p. 7

⁵²[Acar et al., 2014], p. 674

⁵³[Acar et al., 2014], p. 676

⁵⁴https://en.wikipedia.org/wiki/Web_storage, last visited 2021-11-23

⁵⁵[Ayenson et al., 2011], p. 7

⁵⁶[Bujlow et al., 2015], p. 6

⁵⁷<https://support.mozilla.org/en-US/kb/clear-cookies-and-site-data-firefox>, last visited 2021-12-05

⁵⁸[Ayenson et al., 2011], p. 12

⁵⁹[Lerner et al., 2016], p. 1007

Belloro and Mylonas were able to show, that 57,72% of all websites in their data set used the HTML5 localStorage, and of those 70% used it for tracking purposes⁶⁰.

HTML5 IndexedDB:

The HTML5 IndexedDB is the successor of the Web SQL Database and operates under the same condition as the HTML5 localStorage⁶¹. In 2014 Acar et al. were the first to report a new evercookie vector using the HTML5 IndexedDB: They found a script on weibo.com which mirrored a Flash cookie in a user's IndexedDB. This particular Flash cookie was known to respawn HTTP cookies after deletion⁶². Belloro and Mylonas showed in 2018 that, of the 1.68% of websites in their data set which utilised the HTML5 IndexedDB, 30% did it for tracking purposes⁶³. Assuming these numbers reflect the overall popularity of IndexedDB-based web tracking methods, I will not discuss them further in this thesis.

Cache-based tracking mechanisms

ETags:

ETags, or entity tags, are used to identify if resources in a user's web cache are outdated or not⁶⁴ It is a standard part of HTTP and was created to save bandwidth. Its first use for user tracking was found by Ayenson et al. in 2011 on hulu.com, which used ETags to respawn HTTP cookies. In their publication they state: *"ETag tracking and respawning is particularly problematic because the technique generates unique tracking values even where the consumer blocks HTTP, Flash, and HTML5 cookies. In order to block this tracking, the user would have to clear the cache between each website visit. Even in private browsing mode, ETags can track the user during a browser session."*⁶⁵

Nicolas Hinternesch, "a London-based creative digital professional & senior analytics consultant",⁶⁶ showed how ETags can be easily used for user tracking in his article for the online platform "Medium". He was able to track re-visiting users by simply placing an iFrame, sized one by one Pixel, on a website and overwriting its initially generated ETag ID with an user-specific ID. Every time a user returns to the website, their browser cache is now sending the user-specific ETag ID and therefore makes this user identifiable⁶⁷.

Fingerprinting

A fingerprint is something considered uniquely identifying to a person. The same applies to certain combinations of your computer's hard- and software as well as information stored on it. Li and Al-Fannah provide a website, which allows user's to see some of the

⁶⁰[Belloro and Mylonas, 2018], p. 52783

⁶¹[Bujlow et al., 2015], p. 7

⁶²[Acar et al., 2014], p. 681

⁶³[Belloro and Mylonas, 2018], p. 52783

⁶⁴<https://tools.ietf.org/html/rfc7232#section-2.3>, last visited 2020-02-16

⁶⁵[Ayenson et al., 2011], p. 14

⁶⁶<https://hinternesch.com/>, last visited 2020-02-16

⁶⁷<https://levelup.gitconnected.com/no-cookies-no-problem-using-etags-for-user-tracking-3e745544176b>, last visited 2020-02-16

2. LITERATURE REVIEW: USER TRACKING

"fingerprintable" attributes of their setup.⁶⁸ More information about the attributes and type of fingerprints can be found below.

Browser fingerprinting:

Both, Mayer in 2009⁶⁹ and Eckersley in 2010, described their findings on browser fingerprinting. Eckersley's study focused on the version and configuration, which he requested from user's browsers when a certain website was visited.⁷⁰ In his study he found that in a *"sample of privacy-conscious users, 83.6% of the browsers seen had an instantaneously unique fingerprint, and a further 5.3% had an anonymity set of size 2. Among visiting browsers that had either Adobe Flash or a Java Virtual Machine enabled, 94.2% exhibited instantaneously unique fingerprints and a further 4.8% had fingerprints that were seen exactly twice."*⁷¹

In their publication from 2018, Al-Fannah et al. describe browser fingerprinting as a combination of collecting and analysing HTTP request as well as downloading certain JavaScript (JS) files to a user's browser to gather further information.⁷² In fig. 2.7 Al-Fannah et al. list the 10 most collected browser attributes collected by fingerprinters, they also found that *"the most widely used fingerprinting third-party was Google Analytics"*.⁷³ Furthermore, the next 4 domains in the top 5 third-party fingerprinting domains also belong to Google Inc.⁷⁴ In a later paper, which Al-Fannah wrote together with Mitchell, they claimed that Google might not be interested in limiting browser fingerprinting in Google Chrome, being a key player in fingerprinting themselves.⁷⁵

In their 2018 paper, Vastel et al. showed that browser fingerprints change regularly, which complicates long term user tracking but does not make it impossible. They propose a software, called FP-Stalker, to determine whether a new fingerprint should be given a new ID or linked to an existing one, making the fingerprint only an evolution of an already encountered browser instance.⁷⁶ They also provide a web service determining a user's browser uniqueness.⁷⁷

Device fingerprinting / Cross-browser fingerprinting:

Most of the top 10 attributes collected by fingerprinters (see fig. 2.7) do not belong to the browser. Some of them are provided by the Internet Service Provider (ISP), others are giving away information about the system on which the browser is installed.

Boda et al. showed in their 2012 publication that it is possible to identify users who are using more than one browser on the same machine. *"The user ID is the script-generated*

⁶⁸<https://fingerprintable.org/test>, last visited 2021-12-09

⁶⁹[Mayer, 2009]

⁷⁰[Eckersley, 2010], p. 4

⁷¹[Eckersley, 2010], p. 2

⁷²[Al-Fannah et al., 2018], p. 483

⁷³[Al-Fannah et al., 2018], p. 489

⁷⁴[Al-Fannah et al., 2018], p. 491

⁷⁵[Al-Fannah and Mitchell, 2020], p. 173

⁷⁶[Al-Fannah et al., 2018], p. 490

⁷⁷[Vastel et al., 2018], p. 731

⁷⁸<https://amiunique.org/>, last visited 2021-12-19

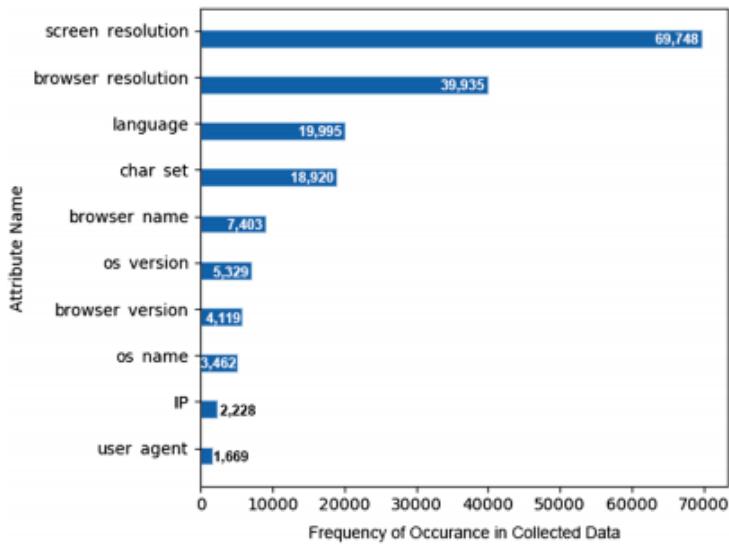


Figure 2.7: Top 10 attributed collected by fingerprinters, from Al-Fannah et al.'s 2018 study⁷⁶

identifier, derived from the first two octets of the IP address, the screen resolution, the time zone, and the 'basic fonts' variables".⁷⁹ By combining these attributes Boda et al. were able to create a fingerprint which was still valid after updates to the computer and/or browser; also (de-)installing plugins or emptying local storage had no effect on its validity.⁸⁰

While Acar et al. distinguished between 4 different types of browser fingerprinting in their 2013 paper (JavaScript-based, plugin-based, extension-based and header-based as well as server-side),⁸¹ Nikiforakis et al. created a taxonomy with 5 fingerprinting categories for their 2013 paper. They distinguish between Browser customisations, Browser-level user configurations, Browser family and version, Operating System and Applications and Hardware and Network.⁸² Some attributes contribute to more than one category. Overall, tracking more attributes allows for a more accurate fingerprint and might even bypass anti-fingerprinting measures. When looking at user-agent-spoofing extensions for Mozilla Firefox and Google Chrome, Nikiforakis et al. found that none of them altered the screen object.⁸³ Meaning, those anti-fingerprinting extensions created impossible configurations and did not hide the most collected attribute from fig. 2.7.

⁷⁹[Boda et al., 2012], p. 35

⁸⁰[Boda et al., 2012], p. 38

⁸¹[Acar et al., 2013], p. 1130 f.

⁸²[Nikiforakis et al., 2013], p. 543

⁸³[Nikiforakis et al., 2013], p. 552

⁸⁴[Acar et al., 2013], p. 1134

2. LITERATURE REVIEW: USER TRACKING

Acar et al. developed a tool to find JavaScript-based font probing scripts on websites.⁸⁵ Figure 2.8 shows a histogram representing the Top 1 Million Alexa sites (the company Alexa Internet publishes global and country-specific ranks based on a website's traffic⁸⁶). Each interval of 100K sites has two bars: the darker one representing the number of websites which were serving fingerprinting scripts known for JavaScript-based font probing; the lighter one representing the number of websites which served and executed those scripts. Acar et al. stated that not all scripts probe a large number of fonts every time they are loaded.⁸⁷

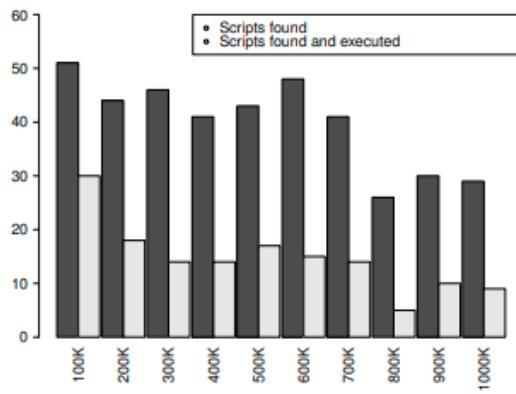


Figure 2.8: JavaScript-based font probing scripts on homepages of Top 1 Million Alexa sites from Acar et al.'s paper⁸⁴

Canvas fingerprinting:

Canvas fingerprinting is a form of browser and device fingerprinting first presented by Mowery and Shacham. It relies on the HTML5 element <canvas>, which "provides an area of the screen which can be drawn upon programmatically"⁸⁸. The drawing process is aided by using a device's graphics processing unit for 2D and 3D graphics and its operating system's font rendering code for text. Therefore, the observed browser's behaviour (e.g. time needed for rendering) gets closely tied to these components and allows for fingerprinting.⁸⁹ Mowery and Shacham were able to distinguish 116 distinct groups of users in their 294 samples, the largest group contained 51 samples. They assumed they would have been able to differentiate between users even further with a few more sophisticated alterations to their setup.⁹⁰

In 2014 Acar et al. did the first study on real-world canvas fingerprinting and found that more than 5.5% of the Top Alexa 100,000 sites actively ran canvas fingerprinting scripts.⁹¹ Most of those scripts could be traced back to a single provider, addthis.com.

IP address tracking:

An Internet Protocol (IP) address is a numerical label, each device connected to a computer network communicating via IP is assigned one of these labels. Although a large portion of IP addresses are dynamic (meaning they are leased to a device only for a period of time and released to be reused when the lease is not renewed), Mishra et al.

⁸⁵[Acar et al., 2013] p. 1131 ff.

⁸⁶<https://www.alexa.com/about>, last visited 2021

⁸⁷[Acar et al., 2013], p. 1135

⁸⁸[Mowery and Shacham, 2012], p. 2

⁸⁹[Mowery and Shacham, 2012], p. 1

⁹⁰[Mowery and Shacham, 2012], p. 8

⁹¹[Acar et al., 2014], p. 678

found that 87% of users had at least one IP address which they retained for more than 30 days.⁹² Thus, making IP addresses a good identifier for user tracking.

Additionally, IP addresses are often handed to third-parties without the user's immediate knowledge or consent, e.g. if a website is using Google Fonts this website submits the user's IP address automatically to Google.⁹³ The website BuiltWith reports 42,781,913 websites using Google Fonts API on the Internet on June 24th 2021,⁹⁴ of which 1,479 are located in Austria.⁹⁵

Other web tracking mechanisms

Web Beacons:

Bouguettaya and Eltoweissy described it as follows in 2003: *"A Web beacon—also known as a Web bug, pixel tag, or clear gif—is a small transparent graphic image that is used in conjunction with cookies to monitor users' actions. A Web beacon is placed in the code of a Web site or a commercial email to let the provider monitor the behavior of Web site visitors or those sending an email. When the HTML code associated with a Web beacon is invoked (to retrieve the image), it can simultaneously transfer information such as the IP address of the computer that retrieved the image, when the Web beacon was viewed, for how long, and so forth."*⁹⁶ One of such web beacons is Google's Analytics pixel.⁹⁷ This method is bypassing certain anti-tracking measures, like prohibiting the transfer of cookies to third parties or blocking JavaScript files from executing.

Supercookies:

Supercookies, or Unique Identifier Header, can be added to the HTTP header of user-generated HTTP requests by the respective ISP to track said user.⁹⁸ Madelyn Bacon stated the following in her definition of supercookies for SearchSecurity: *"Supercookies can be used to collect a wide array of data on users' personal internet browsing habits including the websites users visit and the time they visit them. It does not matter which browser is being used or if users switch browsers. Supercookies can also access information collected by traditional tracking cookies – including login information, cached images and files and plug-in data – and store that information even after the traditional cookie has been deleted. Each supercookie can get as large as 100 KB."*⁹⁹

⁹²[Mishra et al., 2020], p. 809

⁹³https://developers.google.com/fonts/faq#what_does_using_the_google_fonts_api_mean_for_the_privacy_of_my_users, last visited 2021-12-09

⁹⁴<https://trends.builtwith.com/websitelist/Google-Font-API>, last visited 2021-06-24

⁹⁵<https://trends.builtwith.com/websitelist/Google-Font-API/Austria>, last visited 2021-06-24

⁹⁶[Bouguettaya and Eltoweissy, 2003], p. 43

⁹⁷<https://developers.google.com/analytics/resources/concepts/gaConceptsTrackingOverview>, last visited 2022-04-16

⁹⁸[Bujlow et al., 2015], p. 12 & <https://searchsecurity.techtarget.com/definition/supercookie>, last visited 2021-02-15

⁹⁹<https://searchsecurity.techtarget.com/definition/supercookie>, last visited 2021-02-15

2. LITERATURE REVIEW: USER TRACKING

Verizon was fined 1.35 million US-Dollar by the Federal Communications Commission (FCC) for their use of supercookies in 2016.¹⁰⁰ According to former FCC official and current Assistant Professor at Princeton University Jonathan Mayer, Verizon's supercookies have been used by other companies to re-spawn cookies and posed a serious threat to users' privacy.¹⁰¹

History sniffing:

There exist a large number of history sniffing methods; according to Sanchez-Rola et al. they can be split in two categories: CSS-based and time-based. While CSS-based methods rely on the way certain links are displayed, time-based attacks measure the amount of time needed to access certain third-party resources.¹⁰² Sanchez-Rola et al. implemented a time-based history sniffing tool, called Baking Timer, which measured time differences between JavaScript code with and without cookies. Those differences were then used to determine whether cookies had been already present on a certain browser or not. Another difference in loading times allowed them to decide if a browser was logged in to certain websites.¹⁰³

At the moment, there seems to be no research available on the amount of websites using history sniffing methods. However, due to the fact that it can be done without a user's knowledge it should be considered a serious privacy risk.

2.3.2 Email tracking

Email tracking is related to web tracking, especially web beacons (see section 2.3.1 for more details). As Bender et al. described in their 2016 paper, email tracking is made possible by the use of HTML-based emails and several mail transfer agents.¹⁰⁴ The principle is shown in more detail in fig. 2.9. The image, which is referenced for email tracking, is often referred to by the name "tracking pixel".¹⁰⁵

Bender et al. found that "*out of 4,505 e-mails, 1,266 (28%) were in plain-text format, while the remaining 3,239(72%) were HTML-based. [...] The HTML e-mails contained references to 110,080 external images, with an average of 38 external images per e-mail. [...] 2,292 e-mails contained tracking elements, which equated to a ratio of 51% (71%) among all e-mails (HTML e-mails).*"

Although there is a specific section for third-party tracking within this chapter (see section 2.3.5), information leaked to third-parties by email tracking will be discussed

¹⁰⁰<https://www.washingtonpost.com/news/the-switch/wp/2016/03/07/fcc-cracks-down-on-verizons-supercookies/>, last visited 2021-03-02

¹⁰¹<https://www.washingtonpost.com/news/the-switch/wp/2015/11/24/with-this-hire-the-fcc-could-soon-get-tougher-on-privacy-and-security/>, last visited 2020-03-02

¹⁰²[Sanchez-Rola et al., 2020], p. 24:4

¹⁰³[Sanchez-Rola et al., 2020], p. 24:5 ff.

¹⁰⁴[Bender et al., 2016], p. 3

¹⁰⁵[Englehardt et al., 2018], p. 117

¹⁰⁶[Bender et al., 2016], p. 3 & [Fabian et al., 2021] p. 101702-3

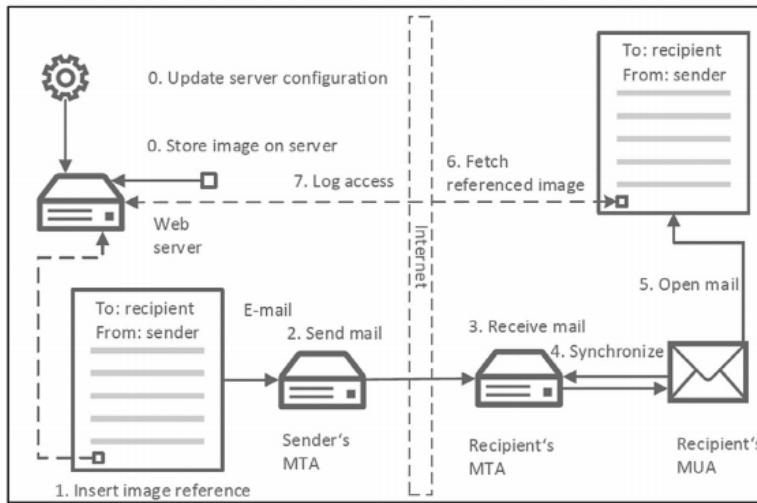


Figure 2.9: The principle of email tracking as described by Bender et al. 2016 and Fabian et al. 2021¹⁰⁶

here. Englehardt et al. observed that at 85% of 12,618 HTML emails (from unique 902 sites) in their data set embedded at least one third-party resource and that many of those third-parties were also involved in web tracking.¹⁰⁷ Englehardt et al. stated that those third-parties are "*blurring the line between email and web tracking.*"¹⁰⁷ They also found, that email tracking content is quite dynamic. Most emails loaded less embedded third-party resources when opened for a second time. However, 21% of emails in the data set loaded third-party content which was not present when the email was opened for the first time.¹⁰⁸

In 2021, Fabian et al. published research on email tracking prevention. Their machine learning approach was able to distinguish tracking from non-tracking images in 99% of the cases.¹⁰⁹

2.3.3 Mobile tracking

To be clear, many of the previously mentioned tracking mechanisms also apply whenever someone uses a mobile browser to access the Internet. A study in 2013 by Eubanks et al. compared web tracking methods employed on desktop and mobile devices. They found that the websites in their data set store more cookies and make more JavaScript calls when accessed from a desktop device.¹¹⁰ However, regarding third-party tracking domains they state: "*In summary, the top trackers on mobile and desktop devices were*

¹⁰⁷[Englehardt et al., 2018], p. 115

¹⁰⁸[Englehardt et al., 2018], p. 116 f.

¹⁰⁹[Fabian et al., 2021], p. 101702-8

¹¹⁰[Eubank et al., 2013], p. 4

2. LITERATURE REVIEW: USER TRACKING

much more similar than we expected."¹¹¹ On top of regular web tracking methods, mobile phones possess their own IDs and a wide range of sensors, which provide completely different tracking angles. Ermakova et al. even found that "*there is a larger privacy threat on mobile phones due to additional privacy-critical information, e.g., end-users' locations, their phone number and contacts, call and email histories, and more*"¹¹². Binns et al. stated that various comparisons show that there are differences between mobile and web tracking with respect to tracking companies as well as information shared with these companies.¹¹³ The following sections contain information on a few prominent examples of mobile tracking methods.

App permissions & third-party tracking in mobile applications

Application permissions, or app permissions, are permissions to access certain (sensor) data given by the user to a specific app on installation or while using the app. The number of permissions and their degree of adjustability vary between operating systems. In 2014, Lin et al. looked into apps' privacy-related behavior, especially into the following 11 app permissions, which they found most sensitive and frequently used: "*INTERNET, READ_PHONE_STATES, ACCESS_COARSE_LOCATION, ACCESS_FINE_LOCATION, CAMERA, GET_ACCOUNTS, SEND_SMS, READ_SMS, RECORD_AUDIO, BLUE_TOOTH and READ_CONTACT*".¹¹⁴ Based on their survey with 725 participants and 21,657 responses regarding 837 apps, they found that there is not a single setting of privacy preferences which will fit all of their participants. However, they came up with the idea of "*segmenting the entire user population into a number of subgroups that have similar preferences within the subgroups*", creating four so-called "privacy profiles".¹¹⁵ By classifying users into one of those "privacy profiles", the number of times those users would have to answer prompts about their privacy-related preferences could be reduced, from every app installation to once with the initial phone setup.

Based on research conducted by Kulyk et al. in 2016, even "*technically-adept people*" have a hard time setting app permissions in a way that could not lead to potential privacy invasion. They came up with four categories of heuristics, see fig. 2.10, for which they developed "*a list of guidelines for supporting users' privacy-related decisions concerning Smartphone apps.*"¹¹⁶ Their guidelines focus on 3 major points: a general risk analysis concept to provide users with the necessary information to estimate privacy risks; a set of recommendations for choosing and installing the right app; and guidelines for managing already installed apps.¹¹⁷

¹¹¹[Eubank et al., 2013], p. 5

¹¹²[Ermakova et al., 2018], p. 4737

¹¹³[Binns et al., 2018], p. 24

¹¹⁴<https://medium.com/swlh/how-mobile-app-permissions-dont-protect-privacy-f749d8fdbfe3>, last visited 2021-06-23

¹¹⁵[Lin et al., 2014], p. 201

¹¹⁶[Lin et al., 2014], p. 203 ff.

¹¹⁷[Kulyk et al., 2016], p. 8

¹¹⁷[Kulyk et al., 2016], p. 8 f.

Even though there is a specific section focusing on third-party tracking (see section 2.3.5), third-party tracking utilised by mobile applications is discussed in the following paragraph for enhanced readability.

App permissions are limited to certain kinds of data, but many mobile applications share other data with third-parties without asking any kind of permission. Zang et al. found that out of their test set of 110 most popular free mobile apps, 73% of Android apps send personal data to third-party domains, while only 16% of iOS apps do the same.¹¹⁸

Additionally to their research on app permissions, Lin et al. looked into the use of third-party libraries in mobile apps. They grouped the libraries of nearly 90,000 (decompiled) apps into 9 categories: *"Targeted Advertising, Customized UI Components, Content Host, Game Engine, Social Network Sites (SNS), Mobile Analytics, Secondary Market, Payment and other Utilities"*. Based on their findings the average app in 2014 used 1.59 third-party libraries, but in extreme cases a single app would use up to 30 third-party libraries.¹¹⁹

In some cases, e.g. WhatsApp sharing personal data with Facebook,¹²⁰ permissions for sharing certain data is buried in the terms of service or privacy policy, leaving people unable to use a service without agreeing to these practices first.

Location-based or on-site tracking

Mobile phones are location-aware, either by built-in sensors, e.g. GPS-sensors, or by networks or devices in their vicinity, sometimes even without being connected to them. Those sensors and connectors are constantly sending out signals, which might be used to track a mobile phone's location without its user being aware of it. Back in 2012, Navizon Inc. claimed *"Unobtrusive surveillance / Navizon I.T.S. works in the background, quietly and unobtrusively locating Wi-Fi- enabled devices... No application is needed on the devices to be tracked. The only requirement is that their Wi-Fi radios be turned on,*

¹¹⁸[Zang et al., 2015]

¹¹⁹[Lin et al., 2014], p. 202

¹²⁰<https://www.forbes.com/sites/carllypage/2021/01/13/whatsapp-clarifies-facebook-data-sharing-as-users-flock-to-rival-signal/>, last visited 2021-06-23

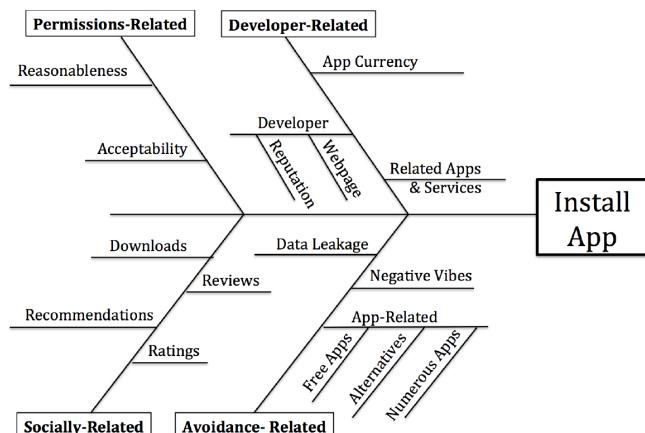


Figure 2.10: Heuristics to decide on app installation (Kulyk et al., 2016)¹¹⁶

2. LITERATURE REVIEW: USER TRACKING

*which is the default in most smart phones, tablets and laptops.*¹²¹ Navizon Inc. has since then moved their indoor tracking division to a company called Accuware,¹²² which still promises their customers they can "get the location of a person or object in a venue in real time".¹²³

The New York Times Privacy Project was able to obtain a file, which contained "50 billion location pings from the phones of more than 12 million Americans". According to the article their data came from software "slipped onto mobile phone apps" by a location data company. As the data was collected over a long period of time, it allowed the journalists to retrieve personal data, e.g. home addresses, based on movement patterns with a high degree of certainty.¹²⁴

Urban et al. state that telecommunication providers routinely store highly accurate location data.¹²⁵ This data has been used in the past, for example by the Austrian telecommunication provider A1 to analyse their customer's movements throughout the first lockdown due to COVID-19.¹²⁶

Enck et. al found that half of their 30 surveyed Android applications sent location data to third-party advertisement servers. In their 2019 paper Boutet and Gambs stated that location is "*is one of the most extensively collected personal data on mobile by applications and third-party services.*"¹²⁷ As mentioned in the section 2.3.3, Lin et al. also listed location access as one of the most sensitive and frequently used app permission. Boutet and Gambs showed in their demonstration the amount of information which can be extrapolated from location data alone. They were able to show where a person lives (together with a picture of that place), their working place (again, with a description and picture), a list of points of interests and personally identifiable information (PII), including probable gender, age and salary.¹²⁸

Device ID (IMEI) & Push tokens

In 2010, Enck et al. wrote the following about device IDs: "[...] the phone contains several easily tainted identifiers: the phone number, SIM card identifiers (IMSI, ICC-ID), and device identifier (IMEI) are all accessed through well-defined APIs."¹²⁹ They surveyed 30 popular third-party Android applications in regard to leaking privacy sensitive information. They found nine of those applications sending IMEIs to their content servers, seven did not inform their users about it.

¹²¹[Urban et al., 2012], p. 5 f.

¹²²<https://www.navizon.com/indoor-positioning-accuware>, last visited 2021-05-18

¹²³<https://accuware.com/>, last visited 2021-05-18

¹²⁴<https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html>, last visited 2021-12-02

¹²⁵[Urban et al., 2012], p. 19

¹²⁶<https://www.derstandard.at/story/2000115828957/mobilfunker-a1-liefert-bewegungsstroeme-von-handynutzern-der-regierung>, last visited 2021-05-18

¹²⁷[Boutet and Gambs, 2019], p. 2861

¹²⁸[Boutet and Gambs, 2019], p. 2863

¹²⁹[Enck et al., 2010], p. 8

IMSI has been quite prominently discussed in the context of IMSI catchers used by Austrian police,¹³⁰ but is not relevant within the scope of this work.

In the last few years another identifier has become more and more relevant within the context of mobile communication, but also throughout the web: push tokens. Those tokens are issued by Apple's or Google's push notification gateways based on a certain app-device combination and needed to send push notifications to a specific device.¹³¹ Due to their identifying nature, push tokens are considered personal data or PII.¹³² Push notifications are not only a vital part of mobile notifications, but can be sent via certain browser features to desktop devices as well. Therefore, notification permission requests have become a staple on many websites, much to the annoyance of website users.¹³³

In 2021, Google implemented an opt-out option for Google Advertising ID and requires all mobile applications published in Google Play Store to use the Google Advertising ID instead of any other device IDs for advertising purposes.¹³⁴ By doing so, they claim to ensure that users can opt-out of targeted in-app advertising. However, it also means the are creating yet another widely used Google-owned UUID. Additionally, data gathered by Singular, a marketing and advertising company, showed that as of June 2021 only about 2% of Android user have turned off ad personalisation.¹³⁵

Apple does also offer an opt-out option via their AppTrackingTransparency framework on devices running iOS 14.5 or iPadOS 14.5 (or any later versions of both). Instead of a general opt-out option, each app has to ask for permission to track.¹³⁶ Their advertisingIdentifier returns only zeros if a user has opted out of being tracked.¹³⁷ Apple's policies also prohibit to fingerprint devices by any other means.¹³⁶

Activity & Health tracking

Data on a person's health is considered sensitive data and especially worthy of protection under the General Data Protection Regulation.¹³⁸ Such information can, however, be easily gathered by current smartphones. Built-in sensors can be used to track mobility patterns, step count, recognise certain kinds of activities, et cetera. As most

¹³⁰<https://www.derstandard.at/story/2000118225114/innenministerium-und-magenta-nennen-zahlen-zur-handyueberwachung-in-oesterreich>, last visited 2021-06-24

¹³¹<https://help.pushwoosh.com/hc/en-us/articles/360000364923-What-is-a-Device-token->, last visited 2021-06-24

¹³²<https://openback.com/product/compliance/>, last visited 2021-06-28

¹³³<https://www.wired.co.uk/article/chrome-firefox-browser-notifications>, last visited 2020-06-24

¹³⁴<https://support.google.com/googleplay/android-developer/answer/6048248>, last visited 2021-12-15

¹³⁵<https://www.singular.net/blog/google-limit-ad-tracking/>, last visited 2021-12-15

¹³⁶<https://developer.apple.com/app-store/user-privacy-and-data-use/>, last visited 2021-12-15

¹³⁷<https://developer.apple.com/documentation/adsupport/asidentifiermanager/1614151-advertisingidentifier>, last visited 2021-12-15

¹³⁸[General Data Protection Regulation, 2016], Rec. 35

2. LITERATURE REVIEW: USER TRACKING

smartphone users carry their phone with them wherever they go, collected data can be extensive. In 2014, Su et al. provide a list of 11 sensor types which were present in most mainstream smartphones and often used for activity recognition.¹³⁹ They also cite a wider range of applications for gathered activity data from a publication by Lockhart et al.: "*Lockhart et al. [Annot.: [Lockhart et al., 2012]] classified the applications of mobile activity recognition according to their targeted beneficial subjects: (1) application for the end users such as fitness tracking, health monitoring, fall detection, behaviour-based context-awareness, home and work automation, and self-managing system; (2) applications for the third parties such as targeted advertising, research platforms for the data collection, corporate management, and accounting; and (3) applications for the crowds and groups such as social networking and activity-based crowd-sourcing.*"¹⁴⁰ Research done by Christl and Spiekermann shows that the second class of applications is already in use. In their section on "Insurance and healthcare", they state that data gathered by digital tracking is used for an insurance's risk assessment.¹⁴¹ As customers with higher risks normally pay higher fees, users should seriously consider what kind of data they provide to their insurance company.

Additionally, according to Su et al.'s paper, a simple accelerometer can be sufficient to create a personal biometric signature.¹⁴² Considering that research from 2013, funded by the California Consumer Protection Foundation and carried out by attorney at law Linda Ackerman, found that "*among the apps with a privacy policy, the majority of technical practices that we considered a risk to users' privacy were not accurately disclosed or described in a way that would enable non-technical users to understand what is actually going on*", activity and health tracking apps should be used with caution.

2.3.4 Cross-device tracking

In 2017 the Federal Trade Commission (FTC) released a report about cross-device tracking. They described the functionality of cross-device tracking as follows: "*Through cross-device tracking, companies can associate multiple devices with the same person.*" In 2016, Vallina-Rodriguez et al. associate a higher privacy risk with cross-platform tracking services because they are able to collect richer behavioural and contextual information about users.¹⁴³ Solomon et al. were the first, to their knowledge, to successfully investigate and measure probabilistic cross-device tracking with 78-96% accuracy in 2018.¹⁴⁴ Their method of cross-platform tracking detection is based on extracted (targeted) ads across platforms.¹⁴⁵

¹³⁹[Su et al., 2014], p. 237

¹⁴⁰[Su et al., 2014], p. 243 f.

¹⁴¹[Christl and Spiekermann, 2016], p. 35

¹⁴²[Su et al., 2014], p. 244

¹⁴³[Vallina-Rodriguez et al., 2016], p. 4

¹⁴⁴[Solomos et al., 2018], p. 2

¹⁴⁵[Solomos et al., 2018], p. 7

The FTC states that cross-device tracking is most relevant to advertisers.¹⁴⁶ A statement which is supported by several analytics/tracking companies promoting their high accuracy in tracking users across multiple devices.¹⁴⁷ For example, Facebook is advertising its "Facebook Pixel"'s cross-device tracking capability first on their respective website.¹⁴⁸ Solomon et al. state, that in sight of their achieved results, they find the high self-reported accuracies by CDT companies believeable.¹⁴⁹

2.3.5 Third-party tracking

Third-party tracking is very common. For example, Sørensen and Kosta found in their 2019 survey of 1,363 websites with 12,778 subpages a total of 3,128 unique third-party domains present.¹⁵⁰ However, only 151 of those were found on a larger number of pages (one or more percent of the data set) and the top 20 were controlled by only a few companies: *"[...] nine TP URLs [Annot.: TP stands for Third-party] controlled by Google [...], two TPs controlled by Facebook [...], Amazon's CDN [...], and the competitor CDN [...], the advertising companies Adnexus [...], criteo.com, adform.net, the analytics companies scorecardresearch.com (TMRG) and gemius.pl, and the omnipresent twitter.com."*¹⁵¹

Third-party cookies

Lerner et al. classified cookie-based third-party trackers into 6 different classes: Analytics, Vanilla, Forced, Referred, Personal and Referred Analytics¹⁵². Third-party trackers can display more than one of these behaviours at the same time. The following list is a direct citation from Lerner et al.'s publication:

1. *Analytics Tracking: [...] Analytics trackers are characterized by a script, sourced from a third party but run in the first-party context, that sets first-party cookies and later leaks those cookies to the third-party domain.*
2. *Vanilla Tracking: The tracker is included as a third party (e.g., an iframe) in the top-level page and uses third-party cookies to track users across sites.*
3. *Forced Tracking: The tracker forces users to visit its domain directly - for example, by opening a popup or redirecting the user to a full-page ad - allowing it to set cookies from a first-party position.*

¹⁴⁶<https://www.ftc.gov/news-events/press-releases/2017/01/ftc-releases-new-report-cross-device-tracking>, last visited 2021-07-08

¹⁴⁷[Solomos et al., 2018], p. 1

¹⁴⁸<https://www.facebook.com/business/learn/facebook-ads-pixel>, last visited 2021-11-25

¹⁴⁹[Solomos et al., 2018], p. 13

¹⁵⁰[Sørensen and Kosta, 2019], p. 1593 f.

¹⁵¹[Sørensen and Kosta, 2019], p. 1595

¹⁵²[Lerner et al., 2016], p. 1001

2. LITERATURE REVIEW: USER TRACKING

4. *Referred Tracking: The tracker relies on another tracker to leak unique identifiers to it, rather than on its own cookies. [...]*
5. *Personal Tracking: The tracker behaves like a Vanilla tracker but is visited by the user directly in other contexts. Personal trackers commonly appear as social widgets (e.g., "Like" or "tweet" buttons)*
6. *Referred Analytics Tracking: Similar to an Analytics tracker, but the domain which sets a first-party cookie is different from the domain to which the first-party cookie is later leaked.*

Third-party cookies are currently starting their fade-out phase (see section 2.2.2 for more details) and might become obsolete in the near future. Until then, they are one of the most common tools to track users across the web.

According to Hu and Sastry it is possible to calculate a "tangle factor". A "tangle factor" is a measurement showing "*how a set of first party websites may be interconnected or tangled with each other based on the common third parties used.*"¹⁵³ They placed all first-party website sharing a third-party in different containers and measured the number of containers necessary to calculate the aforementioned "tangle factor".¹⁵⁴ After calculating the "tangle factor" of the Alexa global Top 500 websites, they tried to reduce the number of containers needed by applying certain anti-tracking measurements; finding that browsers' "Do not track" option did not do a great job (Chrome's option reduced the number of needed containers by three, while Firefox's only reduced it by one). However, Firefox offered an add-on which specifically isolates Facebook logins from other websites. Using this add-on, the number of containers needed fell from 410 to 339. By removing or blocking the top 50 third-party trackers, users with a UK location needed nine containers, while users from China only needed eight.¹⁵⁵

Cookie syncing & Shared identity solutions

Websites sharing an ID with third-parties through cookie syncing can extend their tracking data beyond observed behaviour on their own website and therefore reconstruct a larger fraction of user's browsing patterns.¹⁵⁶ Google calls cookie syncing "cookie matching", it "*allows you to connect first-party data that you own with Google ad data (tracked via Google, DoubleClick, and YouTube IDs) on that same user [...] by combining this data via privacy-centric joins*".¹⁵⁷

¹⁵³[Hu and Sastry, 2020], p. 76

¹⁵⁴[Hu and Sastry, 2020], p. 78

¹⁵⁵[Hu and Sastry, 2020], p. 81 f.

¹⁵⁶[Acar et al., 2014], p. 676

¹⁵⁷<https://developers.google.com/ads-data-hub/guides/cookie-matching>, last visited 2020-02-19

In 2014 Acar et al. have researched cookie synchronisation as part of their publication "The Web Never Forgets: Persistent Tracking Mechanisms in the Wild". They observed the cookie syncing behaviour of the Top 3,000 Alexa domains on Amazon EC2 with and without blocking third-party cookies. Their high level statistics can be seen in fig. 2.11 (the bottom two rows feature minimum/mean/median/maximun). They explain their process as follows: "*We say that an ID is involved in synchronization if it is known by at least two domains. Cookies and domains are involved in synchronization if they contain or know such an ID, respectively. The statistics displayed aggregate both third-party and first-party data, as many domains (e.g. doubleclick.com, facebook.com) exist in both the Alexa Top 3000 and as third-parties on other sites.*"¹⁵⁸

Shared identity solutions make use of cookie syncing, and are trying to solve a larger problem in the digital ad trading ecosystem. As mentioned in section 2.2.2, two major browsers, Mozilla's Firefox and Apple's Safari, are already blocking third-party cookies per default. Google has agreed to do the same for its browser Chrome until 2022, but is replacing third-party cookies with a Google-sponsored alternative at the same time.¹⁵⁹ In their 2020 paper on browser fingerprinting, Al-Fannah and Mitchell stated that Google already planned on doing so in 2014.¹⁶⁰ In March 2021, Google announced they will use a technology called Federated Learning of Cohorts (FLoC) in the future, which will group people with similar interest together and ad tech can only target those cohorts instead of individuals.¹⁶¹ The FLoC White Paper exists publicly since October 2020.¹⁶² In January 2022, Google replaced FLoC due to its controversial nature with another idea called "Topics".¹⁶³

Other first-parties switched to using APIs of so called ID providers for a "third party cookie-less" solution to continue showing users targeted ads. One of most common ways for first parties to include such APIs is to use Prebid.js, a library for header bidding¹⁶⁴ (more information on header bidding can be found in section 3.1.1). Prebid.js supports

¹⁵⁸[Acar et al., 2014], p. 682

¹⁵⁹<https://uk.pc当地/131231/google-effort-to-kill-third-party-cookies-in-chrome-rolls-out-in-april>, last visited 2021-02-25

¹⁶⁰[Al-Fannah and Mitchell, 2020], p. 173

¹⁶¹<https://blog.google/products/ads-commerce/a-more-privacy-first-web/>, last visited 2021-05-18

¹⁶²<https://github.com/google/ads-privacy/blob/master/proposals/FLoC/FLOC-Whitepaper-Google.pdf>, last visited 2021-05-18

¹⁶³<https://techcrunch.com/2022/01/25/google-kills-off-floc-replaces-it-with-topics/>, last visited 2022-04-14

¹⁶⁴[Pachilakis et al., 2019], p. 282

Statistic	Third party cookie policy	
	Allow	Block
# IDs	1308	938
# ID cookies	1482	953
# IDs in sync	435	347
# ID cookies in sync	596	353
# (First*) Parties in sync	(407) 730	(321) 450
# IDs known per party	1/2.0/1/33	1/1.8/1/36
# Parties knowing an ID	2/3.4/2/43	2/2.3/2/22

Figure 2.11: High-level statistics for illustrative crawls under the two third-party cookie settings by Acar et al.¹⁵⁸

2. LITERATURE REVIEW: USER TRACKING

22 different ID providers as of February 2021.¹⁶⁵ Since December 2020, they also provide their own universally unique identifier (UUID): sharedID.¹⁶⁶ "SharedID allows website operators to create an identifier with a standardized format — just a randomized number in a particular format — and store it in their own internet domain as a first-party browser cookie."¹⁶⁷ This ID is then synced "with other demand partners (bidders) competing to win your ad impressions."¹⁶⁸

Other shared identity provider, like Zeotap and LiveRamp, collect login and form field input, e.g. email addresses, and connect those to the data gathered on other clients' websites.¹⁶⁹ Zeotap states on their website that they own "*the world's largest identity graph*", "*the world's largest spine of high-quality identity linkages*" (sourced from "*major telcos and publishers*") and that they are "*the only one to bring third-party identity into the picture*" when it comes to identity resolution.¹⁷⁰

CNAME

In early 2021, Dimova et al. shed light on a tracking method which circumvents certain anti-tracking mechanisms, especially such blocking third-party cookies.¹⁷¹ Canonical Name (CNAME) cloaking makes use of a mechanism where a DNS-resolution does not return an Address record (A record) containing an IP address but a CNAME record containing a reference to another domain name. This process is repeated until an A record is found. "*This means that requests to xxx.example.com may actually be routed to a different site, such as yyy.tracker.com.*"¹⁷² The tracking method is gaining popularity, as fig. 2.12 shows.

Most CNAME trackers use a subdomain of the actual first party, so HTTP requests to their service appear to be same-site requests. By doing so, common anti-tracking mechanisms relying on cookies' SameSite parameter are rendered useless.¹⁷³ When such requests are sent via HTTP instead of a secure connection via HTTPS, they open website visitors up to session-fixation attacks.¹⁷⁴ One CNAME tracker, investigated by Dimova et al., was also vulnerable to cross-site scripting.¹⁷⁵

¹⁶⁵<https://docs.prebid.org/dev-docs/modules/userId.html#bidder-adapter-implementation>, last visited 2021-02-25

¹⁶⁶<https://onlinemarketing.de/programmatic-advertising/prebid-sharedid-launch-publisher-kontrolle>, last visited 2020-02-25

¹⁶⁷<https://prebid.org/product-suite/sharedid/>, last visited 2021-05-18

¹⁶⁸<https://headerbidding.co/sharedid/>, last visited 2021-05-18

¹⁶⁹<https://www.sueddeutsche.de/wirtschaft/cookies-internet-datenschutz-identitaet-1.5479567>, last visited 2021-12-09

¹⁷⁰<https://zeotap.com/platform/identity-resolution/>, last visited 2021-12-09

¹⁷¹[Dimova et al., 2021], p. 2

¹⁷²[Dimova et al., 2021], p. 3

¹⁷³[Dimova et al., 2021], p. 8

¹⁷⁴[Dimova et al., 2021], p. 11

¹⁷⁵[Dimova et al., 2021]

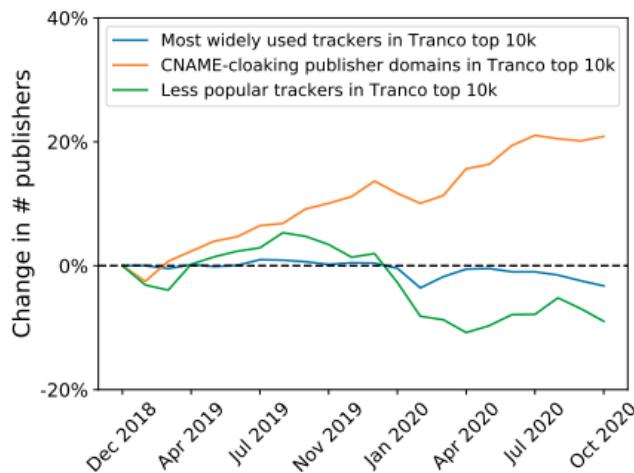


Figure 2.12: Relative percentage, based on the state as of December 2018, of the number of publishers of popular and less popular trackers and CNAME-based trackers from Dimova et al.'s paper¹⁷³

2.3.6 Other user tracking methods

There are quite a few other user tracking methods, for example users can be and are tracked via their wearables, smart TVs or through car telematics. Methods used to do so will not be discussed in the following chapters as all of them are outside of the scope of this work. It is my hope that this work will inspire others to look into them more closely in the future.

3

CHAPTER

Survey: Utilisation of user tracking technologies by Top 500 Austrian Enterprises

3.1 Methodology & Questionnaire design

The second research question of this thesis, as already stated in chapter 1, is:

"Which (state-of-the-art) user tracking technologies are currently used by Austrian enterprises for business purposes?"

In order to answer it, I intended on developing an online questionnaire based on the outcomes of the first research question. However, as Brace already noted in his book about questionnaire design for market research,¹ even the best questionnaire relies on people and what they (can / do not) tell us. So when I realised that few people in my target group would be able to answer very general questions about user tracking methods, I decided to ask about utilization of specific tools. I believed it allowed for a greater sample size. The type of tools chosen were based on the findings from chapter 2, e.g. instead of asking for email tracking, I asked for newsletter tools. How the lists of popular tools were composed is explained in section 3.1.1.

I planned on distributing the questionnaire to various Austrian enterprises to ensure a large enough sample. To determine what size could be considered representative, I looked up numerous guidelines based on best practices and formulas. Many papers, like Hill's "What sample size is “enough” in internet survey research" from 1998,² cite Roscoe's six rules of thumb from 1975.³ The sixth rule, as cited by Hill, is "*There is*

¹[Brace, 2008]

²[Hill, 1998]

³[Roscoe, 1975]

3. SURVEY: UTILISATION OF USER TRACKING TECHNOLOGIES BY TOP 500 AUSTRIAN ENTERPRISES

seldom justification in behavioural research for sample sizes of less than 30 or larger than 500." and Hill assumed it to be applicable to internet surveys.⁴ Another step towards representativity is achieved when a sample matches the characteristic of the targeted population.⁵ The sampled "population" in case of this thesis are Austrian enterprises. Statistik Austria, Austria's Federal Statistical Office, lists 580,393 active Austrian enterprises for the year 2019, of which 341,767 are single-person businesses, 147,892 have 1 to 4 employees, 44,957 have 5 to 9 employees and only 45,777 have 10 or more employees.⁶ A representative sample of Austrian enterprises would therefore contain a large amount of single-person businesses and a considerably smaller amount of businesses with employees. However, it turned out that there is no easy way to contact such a sample. The Austrian Economic Chamber (Wirtschaftskammer Österreich) lists all Austrian enterprises on their website, but does not offer any insight on the size of the regarding enterprise. Statistik Austria only provides aggregated statistics about Austrian enterprises, but no contact data. I tried my best reaching the responsible person at Wirtschaftskammer Österreich, asking them to place a link to my survey in one of their newsletters. However, after more than a month of waiting and several unanswered emails, I had to give up due to time constraints. I resorted to contacting "trend." magazine. Their editorial staff is maintaining a list of Austria's Top 500 enterprises sorted by annual net turnover.⁷ They agreed to provide me with a simplified version of their list for scientific purposes only. A sample of 500 enterprises is still meeting Roscoe's sixth rule of thumb, but the "trend. Top 500" enterprises can hardly be considered representative for all Austrian companies. Their number of employees alone is reason enough to deny their representativity.

In his book, Brace also looked at the impact of wording, order of questions, response categories as well as layout of the questionnaire on the final result. He included a list of problems, that researchers might face when designing a questionnaire.⁸ These were kept in mind while designing the questionnaire for this thesis. Therefore, similar questions are worded comparably, e.g. questions about types of tool are typically starting with "Does your company use..."; lists of tools are always accompanied by the question "Which of the following [x] tools does your company use?", where [x] refers to the type of tool; and all questions are mandatory, but offer two options if a respondent cannot answer: "I do not know" and "I cannot answer this question (e.g. due to legal reasons)". The latter is intended to cover cases where the question might touch areas falling under an employees' non-disclosure agreement.

Papers on surveys done in person as well as those on online surveys agree that a shorter survey length is beneficial to the number and quality of responses. Burchell and Marsh

⁴[Hill, 1998], p. 3

⁵[Biemer, 2010], p. 824

⁶http://www.statistik.at/web_de/statistiken/wirtschaft/unternehmen_arbeitsstaetten/unternehmensdemografie_ab_2015/, last visited 2021-08-06

⁷<https://www.trend.at/wirtschaft/top-oesterreichs-unternehmen-12112369>, last visited 2021-11-24

⁸[Brace, 2008], p. 13 ff.

summarised a number of corresponding literature in their paper from 1992,⁹ before conducting their own experiment with a lengthy interview in person. In 2009, Galesic and Bosnjak looked at the effects of announced and actual length of web surveys on response rate and quality of those responses. They found, that shorter announced lengths led to more started and completed surveys.¹⁰ Regarding the quality of responses, they stated: *"the further away from the beginning of the questionnaire a block of questions was, the less time the respondents spent answering it."*¹¹

With this knowledge in mind and to reduce the number of questions asked to a minimum, I used conditional questions wherever possible. If a respondent answered "No" to using a certain medium or applying a certain type of tracking tools, the whole block of corresponding questions was skipped. It also meant, that the overall survey length varies based on answers given to conditional questions. Therefore, no general survey length was announced in the invitation emails.

The original questionnaire is included in appendix A.2. More details about the implementation and distribution of the actual survey can be found in section 3.2. The following section contains information about how I composed the lists of common user tracking tools, which were offered as possible answers to the respondents, followed by a short section about statistical data on Austrian enterprises.

3.1.1 Common user tracking tools

When not otherwise mentioned, the following criteria were used for sorting the resulting lists: number of mentions in the search results, alphabetical order whenever tools have been mentioned with equal frequency.

Web analytics

The list of web analytics tools, which can be found in appendix A.1, is based on the first page of Google search results for "web analytics tools" (summarised on 2021-09-09). Only results which included a list of more than one tool and have been published or edited later than 2018 were considered.^{12,13,14,15,16,17} Only those tools named by at least two of the results were listed. This measure was taken to reduce bias as more than one of the

⁹[Burchell and Marsh, 1992], p. 233 ff.

¹⁰[Galesic and Bosnjak, 2009], p. 355

¹¹[Galesic and Bosnjak, 2009], p. 356

¹²<https://www.leadfeeder.com/blog/website-analytics-tools/>, last visited 2021-09-09

¹³<https://dynamapper.com/blog/21-sitemaps-and-seo/436-35-amazing-web-analytics-tools-that-rival-google-analytics>, last visited 2021-09-09

¹⁴<https://www.hotjar.com/web-analytics/tools/>, last visited 2021-09-09

¹⁵<https://www.wix.com/blog/2020/01/best-website-analytics-tools/>, last visited 2021-09-09

¹⁶<https://www.trustradius.com/web-analytics>, last visited 2021-09-09

¹⁷https://en.wikipedia.org/wiki/List_of_web_analytics_software, last visited 2021-09-09

3. SURVEY: UTILISATION OF USER TRACKING TECHNOLOGIES BY TOP 500 AUSTRIAN ENTERPRISES

results were part of a curated list on the website of a web analytics tool. 64 results were excluded that way.

One result distinguished between "*traditional analytics tools*" and "*behavior analytics tools*"¹⁴, the difference will be discussed briefly below. The same website also listed three cookieless web analytics tools, which I added to my list even though they were only mentioned once. As I already mentioned in previous sections, cookieless technologies are very likely to become more popular as the use of third-party cookies is getting more and more limited.

Some of the results explicitly tried to name alternatives to Google Analytics, otherwise Google Analytics might have been the most mentioned tool instead of taking second place together with Chartbeat and Mixpanel behind Adobe Analytics. In this regard, Lerner et. al found that in 2011 google-analytics.com covered 35% of all websites in their data set. Therefore, Google would have been able to track users across those websites with Google Analytics (*"via fingerprinting or by changing its behavior to store tracking cookies"*), even though Lerner et al. classified it as "Analytics Tracking" (definition see section 2.3.5).¹⁸

Traditional web analytics:

In 2008, the Digital Analytics Association, formerly the Web Analytics Association, defined web analytics as a "*combination of (a) measuring, (b) acquisition, (c) analyzing and (d) reporting of data collected from the Internet with the aim of understanding and optimizing web experience.*"¹⁹ Traditional tools rely mostly on clickstream data.²⁰ "*Clickstream data are a detailed log of how participants navigate through the Web site during a task. The log typically includes the pages visited, time spent on each page, how they arrived on the page, and where they went next.*"²¹

Behavioural analytics:

Behavioural analytics enhances the limited insights of traditional web analytics with information about a visitor's behaviour. Microsoft states on their website about behavioural analysis: "*Each time a user interacts with your digital channels, they are providing crucial signals about their needs and wants, including readiness to buy.*"²² Capturing those signals, is what behavioural analytics tools do. One example being A/B testing, where distinct groups are served different content and a set of KPIs is compared between groups afterwards. It could also mean that based on your previous behaviour on a website, you get certain ads or offers.²³

¹⁸[Lerner et al., 2016], p. 1008

¹⁹[Bekavac and Garbin Praničević, 2015], p. 374

²⁰[Bekavac and Garbin Praničević, 2015], p. 377

²¹<https://www.sciencedirect.com/topics/computer-science/clickstream-data>, last visited 2021-11-26

²²<https://dynamics.microsoft.com/en-us/ai/customer-insights/what-is-behavioral-analytics/>, last visited 2021-11-26

²³<https://useinsider.com/glossary/behavioral-analytics-2/>, last visited 2021-11-26

Fonts

A web font might not be something that immediately comes to mind in the context of user tracking. However, since 2010 all major browsers support the Cascading Style Sheets (CSS) feature "@font-face".²⁴ Therefore, browsers are able to download and display custom fonts on nearly any device. As mentioned in section 2.3.1, IP addresses are considered personal data and the web font service Google Fonts, for example, is logging IP addresses for all "*of the CSS and the font file requests*".²⁵

Again,a list of common web font directories - which can be found in appendix A.1 - was composed from the first page of Google search results on "web font service"(composed on 2021-09-18), it can be found in appendix A.1. Only results which included a list of more than one tool and have been published or edited later than 2018 were considered.^{26,27,28,29} Only those tools named by at least two of the results are listed. 19 web font services were eliminated from the list that way.

Advertisements

Online advertising is a rather complicated business. There are four main ways for advertisers to place their content on publishers' pages: local or remote ad servers, ad networks, ad exchanges and header bidding. Local ad servers are run by a single publisher and ads are only placed on websites controlled by this publisher. A remote ad server is run by a third-party linking a number of advertisers to a number of publishers. An ad network is closely resembling a remote ad server, but also reduces the overhead for advertisers and publishers. Ad networks handle the complete workflow including choosing ads, reporting and billing.³⁰ A great analogy to understand the difference between ad networks and ad exchanges is provided by MarTech Advisor: *"If an ad network is akin to a stockbroker, then an exchange is like a stock exchange. By serving as an open online marketplace, ad exchanges can do the work of multiple ad networks while also ensuring that everyone has a fair shot at bidding on and winning any ad inventory made available."*³¹ The last layer is header bidding, which was already mentioned in section 2.3.5, where publishers can offer their inventory to more than one ad exchange at the same time. The idea is to increase ad revenue.³² Neither ad exchange tools nor header bidding methods are explicitly

²⁴[Fink, 2010]

²⁵https://developers.google.com/fonts/faq#what_does_using_the_google_fonts_api_mean_for_the_privacy_of_my_users, last visited 2021-09-18

²⁶<https://thoughtbot.com/blog/the-mess-of-web-font-services>, last visited 2021-09-18

²⁷<https://www.hongkiat.com/blog/webfont-comparison/>, last visited 2021-09-18

²⁸https://developer.mozilla.org/en-US/docs/Learn/CSS/Styling_text/Web_fonts, last visited 2021-09-18

²⁹<https://typ.io/libraries>, last visited 2021-09-18

³⁰<https://www.muvi.com/blogs/ad-servers-vs-ad-networks-whats-difference.html>, last visited 2021-11-26

³¹<https://www.inmobi.com/blog/2020/08/10/ad-network-vs-ad-exchange-whats-the-difference-between-the-two>

³²<https://digiday.com/media/wtf-header-bidding/>, last visited 2021-11-21

3. SURVEY: UTILISATION OF USER TRACKING TECHNOLOGIES BY TOP 500 AUSTRIAN ENTERPRISES

mentioned in the survey. Most ad exchange platforms are run by the same companies, sometimes even under the same name as ad networks, e.g. Google Ad Manager. They have therefore been removed from the survey for increased understandability and shorter survey length. Header bidding on the other hand, is mostly done via embedded code snippets and not by simple end-user tools. The technical component of header bidding was considered too advanced for the survey's target audience.

Ad servers:

Like in the previous sections, the list of ad servers - which can be found in appendix A.1 - is based on the first page of Google search results for "ad server" (summarised on 2021-09-20). Only results which included a list of more than one tool and have been published or edited later than 2018 were considered.^{33,34,35} Only those tools named by at least two of the results are listed. This measure was taken to reduce bias as at least one of the results was part of a curated list on the website of an ad serving platform. 9 ad servers were excluded that way. At one occasion DoubleClick for Publishers was changed to Google Ad Manager as Google changed its tool's name in 2018.³⁶ OpenX was removed from the list, as it was decommissioned in 2019.³⁴

Ad networks:

The list of ad networks - which is included in appendix A.1 - is based on the first page of Google search results for "ad networks" (summarised on 2021-09-20). Only results which included a list of more than one tool and have been published or edited later than 2018 were considered.^{37,38,39} Only those tools named by at least two of the results are listed. This measure was taken to reduce bias as at least one of the results was part of a curated list on the website of an ad network. 41 tools were excluded that way. In case of Amazon publisher services the tool's name was changed to "Amazon Ads" to match two other mentions of Amazon's ad network. Verizon Media was removed from the list, the brand got renamed in 2021.⁴⁰

³³<https://clearcode.cc/blog/what-is-an-ad-server/>, last visited 2021-09-20

³⁴<https://www.kevel.co/blog/what-is-an-ad-server/>, last visited 2021-09-20

³⁵<https://www.publift.com/blog/best-ad-servers-for-publishers>, last visited 2021-09-20

³⁶<https://searchengineland.com/google-is-retiring-the-adwords-doubleclick-brands-in-a-major-rebranding-aimed-at-simplification-301073>, last visited 2021-09-21

³⁷<https://www.adpushup.com/blog/the-best-ad-networks-for-publishers/>, last visited 2021-09-20

³⁸<https://www.codefuel.com/blog/best-ad-networks-for-publishers/>, last visited 2021-09-20

³⁹<https://www.singlegrain.com/blog-posts/pay-per-click/alternative-ad-networks/>, last visited 2021-09-20

⁴⁰<https://www.theverge.com/2021/9/2/22653652/yahoo-aol-acquired-apollo-global-management-private-equity>, last visited 2021-12-20

Online shops

Online shops can be embedded in an existing website or run separately. However, it is common, especially for smaller companies, to use existing web shop technologies offered by e-commerce platforms. Some of them have their own (web) analytics tool included. Therefore, they are of interest to this thesis.

Similar to the approach mentioned at the beginning of section 3.1.1, the list - which can be found in appendix A.1 - was composed from the first page of Google search results on "e-commerce platforms" (summarised on 2021-09-16). Only results which included a list of more than one tool and have been published or edited later than 2018 were considered.^{41,42,43,44,45,46} Only those tools named by at least two of the results are listed. As the results were rather coherent, only 9 platforms could be excluded that way.

Newsletter

Again, the list of newsletter tools - included in appendix A.1 - was composed from the first page of Google search results for "newsletter tools" (summarised on 2021-09-21). Only results which included a list of more than one tool and have been published or edited later than 2018 were considered.^{47,48,49,50,51,52,53,54,55} Only those tools named by at least three of the results are listed. This measure was taken to factor in the increased number of search results considered for the list and to reduce bias as more than one of the results were part of a curated list on the website of a newsletter tool. 55 newsletter tools were excluded that way.

⁴¹<https://www.ecommerceceo.com/ecommerce-platforms/>, last visited 2021-09-16

⁴²<https://ecommerce-platforms.com/articles/top-6-ecommerce-platform-reviews-2012-shopify-volusion-bigcommerce-magento-bigcartel-3dcart>, last visited 2021-09-16

⁴³<https://www.websitebuilderexpert.com/ecommerce-website-builders/platforms/>, last visited 2021-09-16

⁴⁴<https://www.techradar.com/news/the-best-ecommerce-platform>, last visited 2021-09-16

⁴⁵<https://influencermarketinghub.com/best-ecommerce-platform/>, last visited 2021-09-16

⁴⁶<https://www.mageplaza.com/blog/ecommerce-platform.html>, last visited 2021-09-16

⁴⁷<https://blog.hubspot.com/marketing/email-newsletter-tools>, last visited 2021-09-21

⁴⁸<https://omr.com/de/beste-newsletter-tools/>, last visited 2021-09-21

⁴⁹<https://www.emailtooltester.com/en/email-marketing-services/>, last visited 2021-09-21

⁵⁰<https://www.blogmojo.de/newsletter-tools/>, last visited 2021-09-21

⁵¹<https://t3n.de/news/e-mail-marketing-anbieter-280807/>, last visited 2021-09-21

⁵²<https://www.omt.de/online-marketing-tools/newsletter-tools/>, last visited 2021-09-21

⁵³<https://zapier.com/learn/email-marketing/best-email-newsletter-software/>, last visited 2021-09-21

⁵⁴<https://www.tooltester.com/de/blog/newsletter-tools/>, last visited 2021-09-21

⁵⁵<https://moosend.com/blog/email-newsletter-software/>, last visited 2021-09-21

3. SURVEY: UTILISATION OF USER TRACKING TECHNOLOGIES BY TOP 500 AUSTRIAN ENTERPRISES

Apps

App Stores:

Both major app distribution platforms, Google's Play Store and Apple's App Store, provide app statistics. Google's app statistics include, for example, metrics on app installations, app sales, ratings and crash reports. Data can be selected for certain dimensions, like Android version, device name, a user's country, a user's wireless carrier or app version.⁵⁶ Apple provides insight on search terms used to find and install an app as well as the impact of marketing campaigns on app sales. Historical data on "*number of installations, sessions, and active devices*" can be used to "*evaluate the impact of product changes*".⁵⁷

Google and Apple also present app providers with the opportunity to target ads via their advertising ID. More information about those can be found in section 2.3.3.

Ads in Apps:

The list of ad networks - which you can find in appendix A.1 - is based on the first page of Google search results for "app ad network" (summarised on 2021-09-22). Only results which included a list of more than one tool and have been published or edited later than 2018 were considered.^{58,59,60,61} Only those tools named by at least two of the results are listed. This measure was taken to reduce bias as at least one of the results was part of a curated list on the website of a mobile ad network. 41 networks were excluded that way. One mention of Facebook was changed to Audience Network to match another reference of Facebook's mobile ad network. Unity was renamed to Unity Ads on one occasion to match another mention.

3.1.2 Statistical data on Austrian enterprises

To understand trends in certain industries and to evaluate which laws are applicable, I included a limited number of questions of statistical nature about the responding company in the survey. The exact questions can be found on the last page of appendix A.2. The answer options for questions with drop-down lists can also be found there, but in the beginning of the section. Those are based on company size and industry classification provided by Statistik Austria, Austria's Federal Statistical Office.^{62,63}

⁵⁶https://support.google.com/googleplay/android-developer/answer/139628?ref_topic=7071935, last visited 2021-11-26

⁵⁷<https://help.apple.com/app-store-connect/#/dev598cef242>, last visited 2021-11-26

⁵⁸<https://www.businessofapps.com/ads/mobile-ad-network/>, last visited 2021-09-22

⁵⁹<https://www.monetizemore.com/best-app-ad-networks/>, last visited 2021-09-22

⁶⁰<https://www.publift.com/blog/best-mobile-ad-networks-for-publishers>, last visited 2021-09-22

⁶¹<https://messapps.com/allcategories/marketing/top-15-mobile-app-ad-networks-and-platforms/>, last visited 2021-09-22

⁶²https://www.statistik.at/web_en/statistics/Economy/trade_services/structural_business_statistics/049989.html, last visited 2021-11-26

⁶³http://www.statistik.at/web_de/statistiken/wirtschaft/unternehmen_arbeitsstaetten/unternehmensdemografie_ab_2015/, last visited 2021-08-06

3.2 Implementation & Distribution

3.2.1 Implementation

Because TU Wien does not host its own survey server, the questionnaire was implemented using the German platform "SoSci Survey".⁶⁴ As the survey does not ask for personal data, I had no issues with complying to the GDPR. However, I still chose SoSci Survey because their servers are hosted in Munich by a company called "PartnerGate GmbH" and there is no data transmitted to countries outside the European Union.⁶⁵ Also, SoSci Survey deletes any data gathered in connection with a survey automatically 94 days after a survey's administrator last logged in. They only keep a list of email recipients who have opted out of receiving information or links to surveys created on their platform. Some of Austria's Top 500 enterprises were already on that list before I tried to contact them (more details in section 3.2.2).

Most questions were designed as single- or multiple-choice questions. Only two questions offered possible answers in form of a drop-down list. The original questionnaire and layout are included in appendix A.2. The flow chart in fig. 3.1 gives an overview of conditional and unconditional questions and their connections.

3.2.2 Distribution

Addressees

As mentioned in section 3.1, I received contact data for Austria's Top 500 enterprises (based on their annual net turnover) from ".trend" magazine. There were a total of 588 email addresses. Unfortunately, the quality of those email addresses was rather poor. In most cases they were scraped from publicly available information on contact or legal imprint pages. Sometimes these addresses belonged to customer care. Quite a number of them returned undeliverable after the first mass email was sent, most of them were replaced by other publicly available addresses of the same company.

Pretest

To avoid major problems with a questionnaire, Brace suggests to at least run an informal pilot of the questionnaire.⁶⁶ SoSci Survey offers a pretest option before sending out invitation links to the actual target group. I sent one of those pretest links to a fellow student at TU Wien and asked him to give me feedback on wording, the questionnaire's flow and possible typos. I incorporated his suggestions regarding both questions about cookies before I moved forward with the survey distribution.

⁶⁴<https://www.soscisurvey.de/en/index>, last visited 2021-11-17

⁶⁵<https://www.soscisurvey.de/en/data-protection>, last visited 2021-11-17

⁶⁶[Brace, 2008], p. 117

3. SURVEY: UTILISATION OF USER TRACKING TECHNOLOGIES BY TOP 500 AUSTRIAN ENTERPRISES

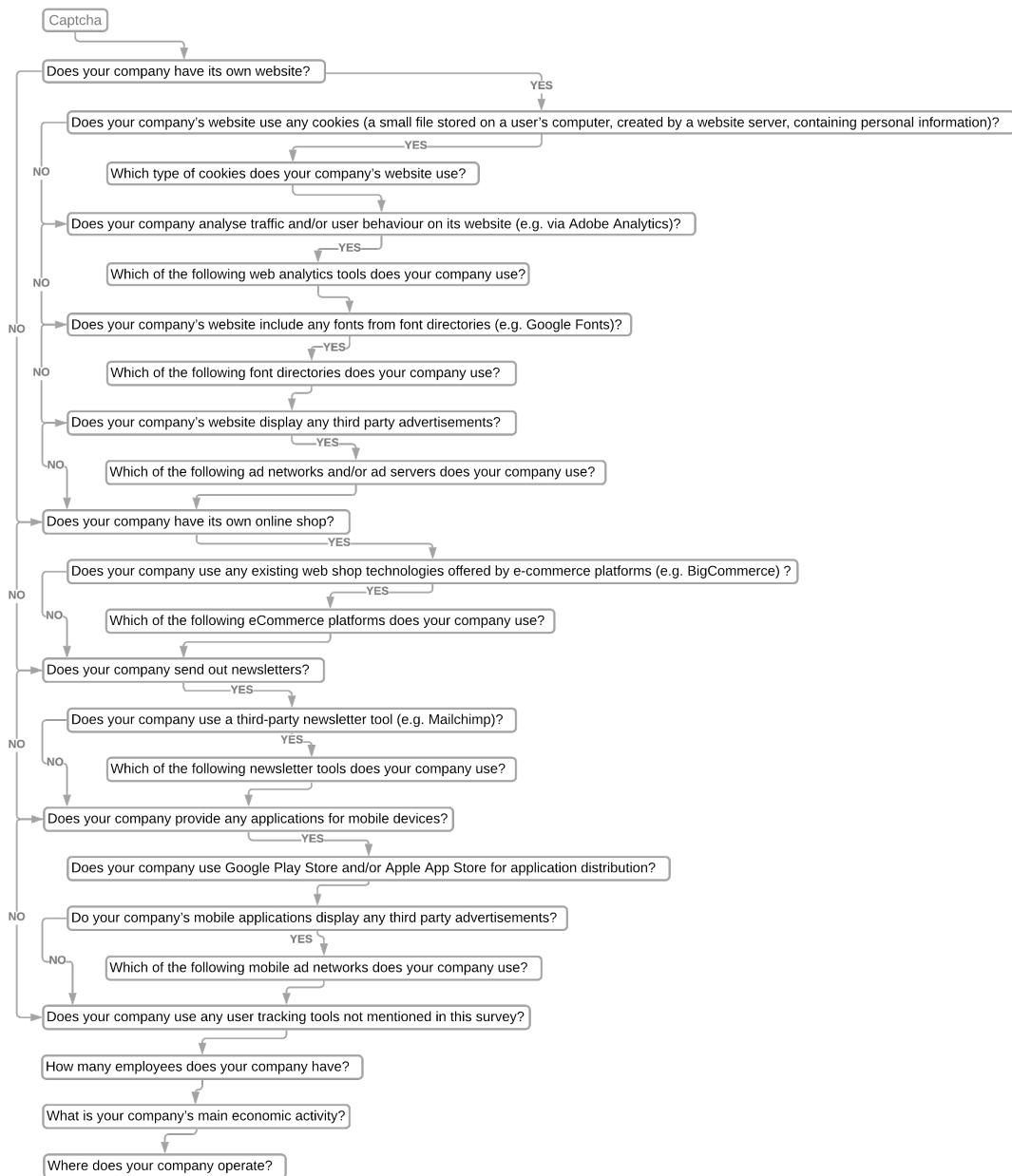


Figure 3.1: Flowchart of the questionnaire

Mass emails

Emails were sent out in two waves, an initial one in early October 2021 and a reminder from end of October to beginning of November 2021. Both waves did not result in a significant amount of responses, partly due to a number of technical difficulties.

The first mass email was sent to 588 addresses, of which 12 were blocked and 4 produced other errors which could not be resolved. A total of 25 surveys were opened, but only a single one was finished by its recipient. After the reminder, which was sent to all addresses with unopened or unfinished surveys, 2 more recipients responded.

After a public complaint about the unsatisfactory turnout, I received 3 more answers from companies working with or related to Austria's Top 500 enterprises.

Some companies sent me an email apologising for not completing my survey. Most of them explained their lack of response by their limited resources, some of them only help students directly working for or associated with them and others have guidelines in place prohibiting them from answering my survey.

Another part of the meager turnout was caused by human error: Due to problems with the mail server, emails got delivered later than expected and contained expired survey links as SoSci Survey features two separate survey link expiration date settings and one of them was missed in the process.

3.3 Results

The following results are only used complementary to the ones in section 4.3. They are neither representative nor statistically meaningful due to the small number of actual responses. However, the scope of the survey included more tracking methods than web scraping can ever reveal. Therefore, the gathered data is used on certain occasions, where information obtained in chapter 4 is insufficient.

Due to the nature of some questions, e.g. text fields for user input, some of the gathered data needed cleansing. As preparation for this thesis, I read Maletic and Marcus' chapter on "Data Cleansing: A Prelude to Knowledge Discovery" published in the "Data Mining and Knowledge Discovery Handbook".⁶⁷ They summarized different error finding and data cleansing strategies. In light of the small number of responses, there was no need to apply such sophisticated methods.

Some data was temporarily lost in the processing step due to limitations of the tools used. SoSci Survey saves the offered fallback options ("I do not know" and "I cannot answer this question (e.g. due to legal reasons)") as negative numbers, which were removed from the downloaded response data as soon as it was opened with Microsoft Excel. The limited number of responses allowed for a quick manual recovery of missing data.

⁶⁷[Maletic and Marcus, 2010]

3. SURVEY: UTILISATION OF USER TRACKING TECHNOLOGIES BY TOP 500 AUSTRIAN ENTERPRISES

3.3.1 Response analysis

100% of responding companies have their own website and all of them store cookies in their visitors' browsers. Interestingly, two out of six companies do not use first-party cookies, but all companies answered yes when asked if they use third-party cookies. Based on the questions' wording, these answers might also include cookies set by a third-party tool under the first party's domain. All companies apply web analytics tools to their websites, some even more than one. The analytics tool used most is Google Analytics, which is used by 83.34% of the responding companies. Other analytics tools used are Adobe Analytics, AT Internet: Web Analytics, Hubspot, SEMrush and Seobility. Two companies make use of Googles font directory, one company even applies a second library (Font Awesome). Another company chose not to answer this question. Only a single company shows third-party advertisements on their website, but the person answering the survey was not able to give any details on the tool used for this purpose.

Half of the responding companies have their own online shops. One of them uses no e-commerce platform, while another uses three simultaneously (Shopify, WooCommerce and SAP Hybris). The third company which answered yes when asked about owning a online shop, chose not to disclose whether they use an e-commerce platform for it.

Four out of six companies send out newsletters, only one of those does not do it via a third-party newsletter tool. Interestingly, none of the tools used were covered by the list of common newsletter tools composed in section 3.1.1. The responding companies use ELAINE by artegic, Emarsys and mailworx by eworx.

Both companies which are providing mobile applications distribute those apps via Google's Play Store and Apple's App Store. One of them displays third-party advertisements in their app(s), but the person answering the survey did not know which tool was used for this purpose. One respondent did not know if their company provided any mobile applications at all.

The question allowing respondents to enter further tracking tools not mentioned in the survey showed that the questionnaire lacked a question about marketing solutions provided by social media companies. Several companies named Facebook Pixel, which was also found on several pages by the web scraping process explained in chapter 4. Other answers included Firebase Crashlytics, LinkedIn Insight Tag, Permutive and Braze.

3.3.2 Statistical company data

The following results are based on the responses of each company to a small number of not identifying statistical questions.

All companies, except for one, have 250 or more employees. This was to be expected as Austria's Top 500 enterprises are either large companies themselves, or a branch or part of larger groups. Out of the 6 answering companies, the main economic activity of two of the responding companies is "Wholesale and retail trade; repair of motor vehicles and motorcycles", one focuses on "Information and communication", another on

"Manufacturing". The remaining two companies answered the regarding question with "Other service activities". In the last question the respondents were asked where their companies operate. It was self-evident they all operate in Austria, however, four out of six also operate in other countries within the European Union and two of those operate outside the EU as well.

CHAPTER 4

Web scraping: Utilisation of user tracking technologies by Top 500 Austrian Enterprises

4.1 Methodology & Definition

To enhance the limited data gathered from the small number of answered surveys, I opted for another data retrieval method: Web scraping. Unfortunately, not all user tracking methods mentioned in section 2.3 can be captured by a simple web scraper.

4.1.1 Definition of web scraping

In his book on the subject, Mitchell defined web scraping, in theory, as *the practice of gathering data through any means other than a program interacting with an API (or, obviously, through a human using a web browser)*.¹ In practice, web scraping is a form of data mining, extracting certain data from websites. Web scraping is targeted at specific data on specific pages, while web crawling is scraping for any data available on the Internet.²

4.1.2 Limitations

A number of user tracking methods can be revealed through inspection of a website's cookies, loaded JavaScript files, as well as certain HTTP requests. Results, which can be found in section 4.3, could be impacted by anti-scraping technologies. No special effort was made to detect or bypass those in the self-implemented scripts.

¹[Mitchell, 2018], p. 1

²<https://www.parsehub.com/blog/web-scraping-vs-web-crawling/>, last visited 2021-11-26xsy

4. WEB SCRAPING: UTILISATION OF USER TRACKING TECHNOLOGIES BY TOP 500 AUSTRIAN ENTERPRISES

4.1.3 Browser extensions

For the purpose of this thesis, it would have been ideal to consent to all cookies on browsed websites. However, due to the intricacies of and lack of standardisation in consent notices, implementation of appropriate functionality in my scraping tool was not reasonably possible within the usual time-frame of a master thesis. However, there exist browser extension to deal with most cookie consent banners. More information on these extensions and how they work can be found below.

Consent-O-Matic

An extension developed by privacy researchers at Aarhus University in Denmark.³ It attempts to auto-fill consent notices based on user preferences. It only works on a limited number of consent management platforms (CMPs).⁴ The extension would have been a great option to accept all cookies as it offers cookie preference settings. However, I was unable to find a way to access those settings from within my script(s). In the end, I had to limit myself to the default settings, which do not accept any not strictly necessary cookies.

I don't care about cookies

This extension mostly *"just blocks or hides cookie related pop-ups. When it's needed for the website to work properly, it will automatically accept the cookie policy for you (sometimes it will accept all and sometimes only necessary cookie categories, depending on what's easier to do)."*⁵ Therefore, a slight increase in gathered cookies can be expected when this extension is used.

Ninja Cookie

The "Ninja Cookie" extension takes a different approach and actively tries to accept as little cookies as possible. *"Ninja Cookie is a browser extension that automatically removes cookie banners by rejecting the use of non-essential cookies."*⁶ Due to technical difficulties scraping runs utilising this extensions were not successful and produced no suitable results. Therefore, those runs were excluded from the analysis in section 4.3.

4.1.4 webXray

The scripts shown in section 4.2 only account for first-party data when looking at cookies and first- and third-party data when considering script sources. However, as discussed in section 2.3.5, third-party cookies are still in use and therefore should be part of

³<https://chrome.google.com/webstore/detail/consent-o-matic/mdiildafknihdffpkfmmppnpoiajfjnjd/>, last visited 2021-11-21

⁴<https://github.com/cavi-au/Consent-O-Matic#compatible-cmps>, last visited 2021-11-21

⁵<https://www.i-dont-care-about-cookies.eu/>, last visited 2021-11-17

⁶<https://ninja-cookie.com/>, last visited 2021-11-21

the analysed data as well. To achieve this goal and to ease the matching of gathered JS-files to known third parties, I used a tool called webXray. It is developed by Timothy Libert, a privacy engineer and former faculty member in the School of Computer Science at Carnegie Mellon University.⁷ webXray is meant for "*analyzing webpage traffic and content, extracting legal policies, and identifying the companies which collect user data.*"⁸ Libert gave a detailed explanation of the functionality of webXray in his paper for the International Journal of Communication.⁹ The tool takes a list of web addresses as input. I used the same list as for the self-implemented scripts. Each page is then processed for cookies, HTTP requests and HTTP received events. All gathered data is stored in a database. WebXray provides a number of automatically created reports based on the gathered data, which will be discussed in section 4.3.3.

4.2 Implementation & Execution

4.2.1 Data & data clean up

The data set fed to the web scraper was extracted from the Excel sheet provided by ".trend" magazine (see section 3.2.2 for more information). It includes websites of or associated with Austria's Top 500 enterprises. This means not only websites of said companies but also websites of brands belonging to these companies were included in the data set. After duplicates were removed and rows which included multiple URLs were separated the data set consisted of 618 unique URLs.

4.2.2 Code

The following scripts were run using Python 3.8.1, Selenium 4.0.0 and Chrome Version 95.0.4638.69 with ChromeDriver 95.0.4638.69 on a Windows 10 Enterprise installation. The extensions mentioned in section 4.2.2 were downloaded using a web service provided by CRX extractor.¹⁰

Without extensions

The code in listing 4.1 browses each URL contained in 'urls.csv', waits for the page to load, collects all first-party cookies present via Selenium's `get_cookies()` function and writes name and value of each cookie together with the respective URL to a comma-separated values (CSV) file.

The code in listing 4.2 browses each URL contained in 'urls.csv', waits for the page to load, collects all scripts present via Selenium's `find_elements()` function. Scripts are detected by their HTML tag name and their source together with the respective URL is

⁷<https://timlibert.me/>, last visited 2021-11-23

⁸<https://github.com/timlib/webXray>, last visited 2021-11-22

⁹[Libert, 2015], p. 3550 f.

¹⁰<https://crxextractor.com/>, last visited 2021-11-13

4. WEB SCRAPING: UTILISATION OF USER TRACKING TECHNOLOGIES BY TOP 500 AUSTRIAN ENTERPRISES

written to a file. Inline scripts are skipped as their source is empty and analysing their content would exceed the scope of this thesis.

```
1 import time
2 import csv
3 from selenium import webdriver
4
5 browser = webdriver.Chrome()
6 browser.set_window_size(1024, 768)
7
8 with open('cookies_before_consent.csv', mode='w', newline='') as cFile:
9     cWriter = csv.writer(cFile)
10
11     with open('urls.csv', newline='') as urlFile:
12         urlReader = csv.reader(urlFile)
13
14         for row in urlReader:
15             try:
16                 browser.get("https://" + row[0])
17                 time.sleep(2)
18
19                 cookies = browser.get_cookies()
20                 for cookie in cookies:
21                     cName = cookie['name']
22                     cVal = cookie['value']
23                     if not cName:
24                         continue
25                     cWriter.writerow([row[0], cName, cVal])
26
27             except Exception as e:
28                 cWriter.writerow([row[0], "An error occurred."])
29
30 browser.quit()
```

Listing 4.1: Python script to scrape for cookies

With extensions

Listing 4.3 contains a code snippet used to create a Chrome instance featuring a Chrome extension file (CRX). It extends the scripts of listing 4.1 and listing 4.2. As mentioned at the beginning of this section, the necessary CRX-files were extracted from their respective Chrome Web Store page^{11,12,13} using CRX extractor.¹⁴

```

1 import time
2 import csv
3 from selenium import webdriver
4 from selenium.webdriver.common.by import By
5
6 browser = webdriver.Chrome()
7 browser.set_window_size(1024, 768)
8
9 with open('js_before_consent.csv', mode='w', newline='') as jsFile:
10     jsWriter = csv.writer(jsFile)
11
12     with open('urls.csv', newline='') as urlFile:
13         urlReader = csv.reader(urlFile)
14
15     for row in urlReader:
16         try:
17             browser.get("https://" + row[0])
18             time.sleep(2)
19
20             scripts = browser.find_elements(By.TAG_NAME, 'script')
21             for script in scripts:
22                 src = script.get_attribute('src')
23                 if not src:
24                     continue
25                 jsWriter.writerow([row[0], src])
26
27         except Exception as e:
28             jsWriter.writerow([row[0], "An error occurred."])
29
30 browser.quit()

```

Listing 4.2: Python script to scrape for script sources

¹¹<https://chrome.google.com/webstore/detail/i-dont-care-about-cookies/fihnnjjcciajhdojfnbddfaoknhalnja>, last visited 2021-11-17

¹²<https://chrome.google.com/webstore/detail/consent-o-matic/mdjildafknihdffpkfmmpnpoiajfjnjd/>, last visited 2021-11-21

¹³<https://chrome.google.com/webstore/detail/ninja-cookie/jifeafcpcjjgnlcnkffmeegehmnmkefl>, last visited 2021-11-17

¹⁴<https://crxextractor.com/>, last visited 2021-11-13

```
1 [...]  
2 from selenium.webdriver.chrome.options import Options  
3  
4 op = Options()  
5 op.add_extension('extension_3_3_4_0.crx')  
6  
7 browser = webdriver.Chrome(options=op)  
8 [...]
```

Listing 4.3: Code snippet to use a Chrome extension with Selenium and ChromeDriver

4.2.3 Data scraping

All scripts were run on November 4th and 5th 2021, except for the scripts with Consent-O-Matic extension. Those were executed on November 21st and 22nd 2021. Gathered data is analysed in the following section.

4.3 Results

Most results were analysed using Microsoft Excel. As information gathered by webXray is stored in a SQLite database, SQLite Browser¹⁵ was used for data exploration whenever webXrays automatically generated reports did not offer enough detail.

4.3.1 First-party cookies

On page load

The web scraper, being fed 618 URLs, gathered 2192 first-party cookies on 440 unique websites into one CSV-file. Of those, 1015 were unique by name. 7 rows within the file turned out to be exact duplicates and were removed. 19 websites could not be reached or produced error messages for other reasons. Rows containing error messages were removed.

There are 24 first-party cookies with 10 or more occurrences on websites provided by Austria's Top 500 enterprises. They can be found in fig. 4.1. These cookies were present when the website was browsed with a clean browser instance on page load and no cookie consent has been given.

The Top 24 first-party cookies include a number of obvious tracking cookies: As fig. 4.1 shows, "_ga" and "_gid" were by far the cookies most often reported by the web scraper. With 116 occurrences (_ga) and 113 occurrences (_gid) these cookies are present on 18.8% and 18.2%, respectively, of all pages in the data set. Both are placed by Google

¹⁵<https://sqlitebrowser.org/>, last visited 2021-11-23

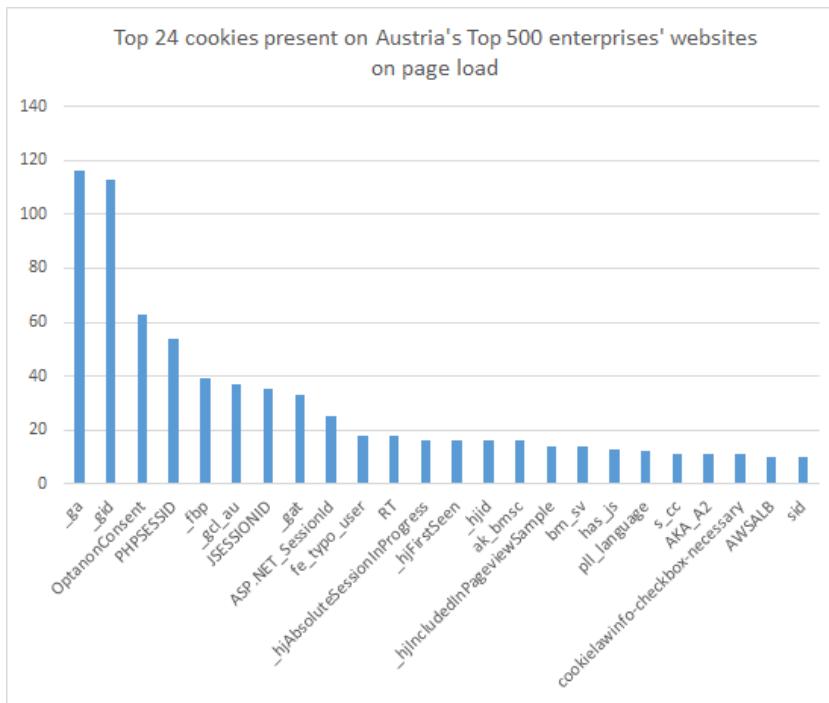


Figure 4.1: The Top 24 first-party cookies, which were found most on the websites of Austria's Top 500 enterprises on page load

Analytics' gtag.js and/or analytics.js, they are "*used to distinguish users*".¹⁶ There are more of Google's cookies present in the 25 most found cookies: There is no information from Google on which Google tool sets "_gcl_au", however, it seems to be linked to Google AdSense;¹⁷ "_gat" is set by Google Analytics;¹⁶ "sid" contains "*digitally signed and encrypted records of a user's Google Account ID and most recent sign-in time*".¹⁸ Google states it uses "sid" for security reasons, other sources state that it is also used in ad optimisation and for retargeting.¹⁹

"_fbp" is a cookie related to Facebook pixel: "*When the Facebook pixel is installed on a website, and the pixel uses first-party cookies, the pixel automatically saves a unique identifier to an _fbp cookie for the website domain if one does not already exist.*"²⁰ It was found on 6.3% of all websites in the data set.

There are four cookies associated with the web analytics tool "HotJar" in the Top 24 cookie, whereof one is meant as a concrete tracking cookie ("_hjid") and the other three

¹⁶<https://developers.google.com/analytics/devguides/collection/analyticsjs/cookie-usage>, last visited 2021-11-13

¹⁷https://cookiedatabase.org/cookie/google-adsense/_gcl_au/, last visited 2021-11-13

¹⁸<https://policies.google.com/technologies/cookies>, last visited 2021-11-13

¹⁹<https://cookiedatabase.org/cookie/google-ads-optimization/sid/>, last visited 2021-11-13

²⁰<https://developers.facebook.com/docs/marketing-api/conversions-api/parameters/fbp-and-fbc/>, last visited 2021-11-13

4. WEB SCRAPING: UTILISATION OF USER TRACKING TECHNOLOGIES BY TOP 500 AUSTRIAN ENTERPRISES

are related but do not contain personal data but simply boolean values ("__hjAbsoluteSessionInProgress", "__hjFirstSeen" and "__hjIncludedInPageviewSample").²¹

The "s_cc" cookie is part of Adobe Analytics, but does not save any personal data. Other cookies set by Adobe's analytics software were all present in the data set, but did not rank in the top 24 cookies, due to their not being consistently set across many websites. Only one page set "s_ecid", another set "s_vi" and "s_fid", the latter was also set by a third website. All of these cookies are used by Adobe Analytics to identify unique visitors.²²

Most of the above mentioned cookies fall under the category of Analytics tracking as defined by Lerner et al.²³, meaning they are cookies set in the first-party context but later leaked to third parties.

Some cookies are not (directly) meant for tracking purposes, but could be used as such with different tracking methods mentioned in section 2.3: "PHPSESSID" (PHP), as well as "JSESSIONID" (J2EE), "ASP.NET_SessionId" (ASP.NET) and "fe_typo_user" (TYPO3), are used to keep track of a user's session by various web frameworks.. "RT" is used to measure a page's loading time. The plugin provider Akamai states that the cookie *"doesn't contain personal information but it contains various pieces of information about the visitor's session, such as number of visited pages, session start time, last visited url and etc."*²⁴ There are other Akamai cookies present in the Top 24 cookies: "ak_bmsc" and "bm_sv". They are used to distinguish between humans and bots according to information on Cookiepedia,²⁵ which Akamai's cookie preferences page directly links to.²⁶ "AKA_A2" is also associated with Akamai and *"used for the Advanced Acceleration feature, which enables DNS Prefetch and HTTP2 Push."*²⁷

Amazon Web Services Load Balancer sets the "AWSALB" cookie, a sticky session cookie, to route requests for a particular session to the same physical machine that serviced the first request.²⁸

Other prominent cookies are not linked to tracking, they do however point out popular tools: The third cookie in the list is "OptanonConsent" and is used for a cookie compliance solution provided by the company OneTrust.²⁹

"has_js" is not solely but mostly set by the Drupal content management system. It saves

²¹<https://help.hotjar.com/hc/en-us/articles/115011789248-Hotjar-Cookie-Information>, last visited 2021-11-13

²²<https://experienceleague.adobe.com/docs/core-services/interface/administration/ec-cookies/cookies-analytics.html>, last visited 2021-11-14

²³[Lerner et al., 2016], p. 1001

²⁴<https://developer.akamai.com/tools-boomerang/docs/tutorial-howto-opt-out-or-opt-in.html#cookies-and-local-storage>, last visited 2021-11-13

²⁵https://cookiepedia.co.uk/cookies/ak_bmsc, last visited 2021-11-13

²⁶<https://www.akamai.com/legal/privacy-and-policies/manage-cookie-preferences>, last visited 2021-11-13

²⁷https://cookiepedia.co.uk/cookies/AKA_A2, last visited 2021-11-14

²⁸<https://docs.aws.amazon.com/elasticloadbalancing/latest/application/sticky-sessions.html>, last visited 2021-11-14

²⁹<https://cookiepedia.co.uk/cookies/OptanonConsent>, last visited 2021-11-13

information about a user's browser and its JavaScript capabilities.³⁰

"pll_language" is set by Polylang, a plugin for internationalisation of WordPress websites, and contains the language code of the last browsed website.³¹

"cookielawinfo-checkbox-necessary" is a cookie placed by the GDPR Cookie Consent plugin provided by Cookie Law Info.³²

The website with the most cookies present before a user has even given consent to cookies being placed, is "www.gehealthcare.com". With 44 unique cookies present it sets 14 more cookies than the website with the second most cookies, "www.sky.at", and 15 more than "www.swarovskigroup.com" and "at.ingrammicro.com". A closer look at those 44 cookies reveal a total of 6 analytics tools, 4 tools related to advertising / targeting and one survey tool. Only 4 cookies could not be assigned to known tools.

The 20 URLs with the most cookies present on page load can be found in fig. 4.4a.

With Consent-O-Matic extension

Executing the web scraping script with Consent-O-Matic extension, which was fed the same 618 URLs as the script without extensions, led to 2329 gathered cookies on 449 unique websites. Of those, 1064 were unique by name. 8 rows within the file turned out to be exact duplicates and were removed. 17 websites were unreachable or produced other errors while scraping, rows produced by those were removed as well.

When the web scraper script is executed featuring the Consent-O-Matic extension, there are 25 first-party cookies with 10 or more occurrences on websites provided by Austria's Top 500 enterprises. Figure 4.2 provides an overview. These cookies were present when the website was browsed with a clean browser instance. Even though the default settings of Consent-O-Matic only accept necessary cookies, the number of gathered tracking cookies increased slightly. The Google Analytics cookies "_ga" and "_gid" could now be found on 20.3% and 19.9%, respectively, of all websites in the data set. There are only two cookies which were not already present in the Top 24 cookies found on page load without extension. The first is "OptanonAlertBoxClosed". It is set by the CMP OneTrust when users actively close the consent pop-up. The second is "CookieConsent". This cookie is used to store a user's consent information and is set by a tool called "Cookiebot". Both tools are listed in the supported CMPs of Consent-O-Matic.³³

HotJar's UUID cookie "_hjid", which was consistently found by other script runs, is not contained in the data gathered by this script at all. As this particular script was executed about 2 weeks after the others, the changes might be due to adjustments in

³⁰https://cookiepedia.co.uk/cookies/has_js, last visited 2021-11-14

³¹<https://polylang.pro/doc/is-polylang-compatible-with-the-eu-cookie-law/>, last visited 2021-11-14

³²<https://cookiedatabase.org/cookie/gdpr-cookie-consent/cookielawinfo-checkbox-necessary/>, last visited 2021-11-14

³³<https://github.com/cavi-au/Consent-O-Matic#compatible-cmps>, last visited 2021-11-21

4. WEB SCRAPING: UTILISATION OF USER TRACKING TECHNOLOGIES BY TOP 500 AUSTRIAN ENTERPRISES

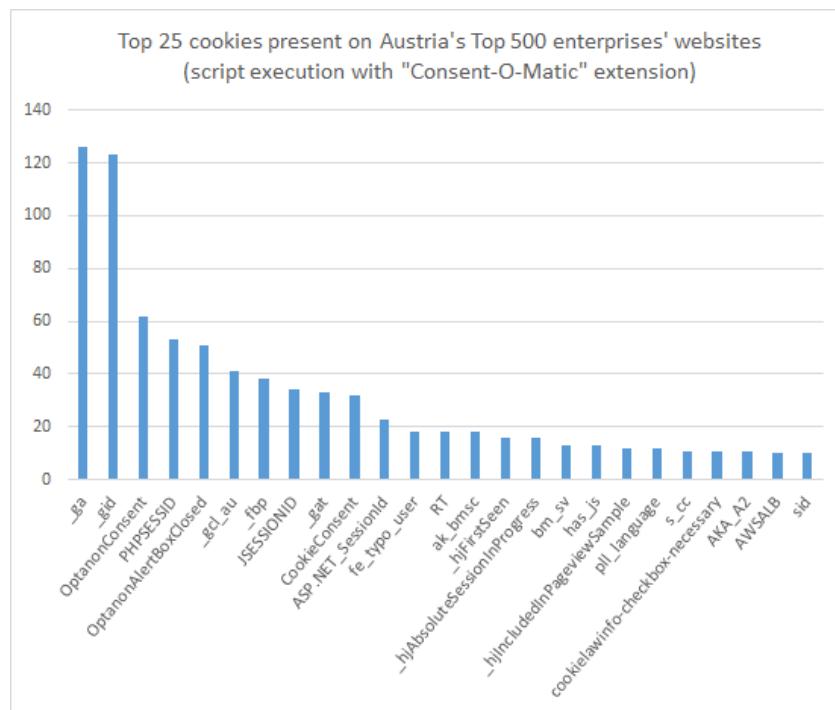


Figure 4.2: The Top 25 first-party cookies, which were found most on the websites of Austria's Top 500 enterprises when browsing with Consent-O-Matic extension

HotJar's software. However, such changes have not yet been publicly announced on their website.³⁴

The differences in the Top 20 websites, based on the number of cookies present, are rather small. Figure 4.4b shows that only two websites were not already featured in fig. 4.4a: "www.egger.com" and "www.pantheon.com". The first, now setting 25 instead of 7 cookies, includes the aforementioned tool Cookiebot. It therefore got permission to set strictly necessary cookies. Their definition of "strictly necessary" includes cookies by Google Analytics, Facebook pixel, HotJar and Jentis (the *#1 Hybrid Tracking Solution for 1st Party Data*).³⁵ Another interesting case is "www.upm.com", which is not part of the Top 20 websites anymore when Consent-O-Matic is used. It seems the presence of the Consent-O-Matic extension leads to a removal of Facebook pixel's cookie "_fbp" and all cookies associated with HotJar. Other cookies, including those set by Google Analytics, remain.

³⁴<https://help.hotjar.com/hc/en-us/articles/115011789248-Hotjar-Cookie-Information>, last visited 2021-11-22

³⁵<https://www.jentis.com/>, last visited 2021-11-22

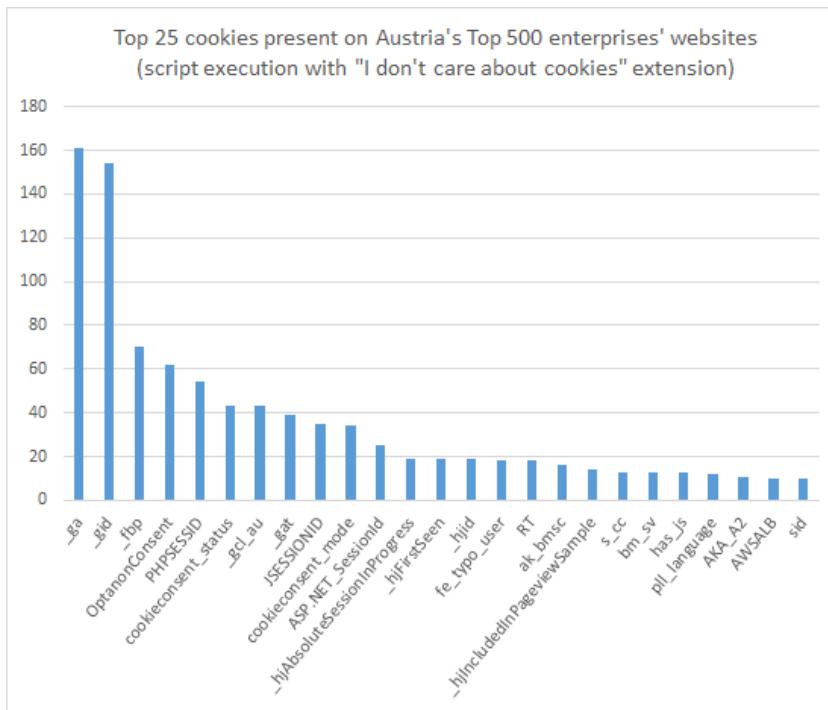


Figure 4.3: The Top 25 first-party cookies, which were found most on the websites of Austria's Top 500 enterprises when browsing with "I don't care about cookies" extension

With "I don't care about cookies" extension

The web scraping script with "I don't care about cookies" extension was fed the same 618 URLs as the one without extension. It gathered 2482 cookies on 458 unique websites into one CSV-file. 1069 of those cookies were unique by name. 7 rows within that file turned out to be exact duplicates and were removed. 19 rows contained error messages produced by websites, which were unreachable or produced other errors. Those were removed from the CSV-file.

There is a significant increase of Google Analytics as well as Facebook pixel cookies gathered when the web scraping script is executed featuring the "I don't care about cookies" extension. "__ga" can now be found on 26.1% of all 618 websites in the data set (without extension: 18.8%); "__gid" is present on 24.9% of those pages (without extension: 18.2%). Facebook's "__fbp" cookie was present on 70 websites, which is an increase of nearly 180% compared to data gathered by the script without extensions.

There are two new cookies, which have not been present in the Top 24 or Top 25, respectively, of the previous discussed script runs: "cookieconsent_status" and "cookieconsent_mode". Both are being set by a CMP called "DataReporter".³⁶ As there is no conclusive information which CMPs are interacting properly with the "I don't care about

³⁶<https://datareporter.eu/de/>, last visited 2021-11-22

4. WEB SCRAPING: UTILISATION OF USER TRACKING TECHNOLOGIES BY TOP 500 AUSTRIAN ENTERPRISES

"cookies" extension, it is assumed that DataReporter does not allow the consent pop-up to be hidden and cookies are therefore accepted.

The "cookielaawinfo-checkbox-necessary" cookie, which was present in the Top 24 / Top 25 of the script execution without extensions and the one with Consent-O-Matic extension, is not listed in the Top 25 of this script run. Cookie Law Info's CMP seems to be blocked on most websites. On the only two websites where such cookies were present, the extension accepted necessary cookies twice, non-necessary cookies once and denied consent for advertisement and analytics cookies.

Two websites rank significantly higher in fig. 4.4c than in fig. 4.4a: "www.teufelberger.com" and "www.peek-cloppenburg.at". Two websites are new to the Top 20 overall: "www.sap.at" and "www.mazda.at". All of it can be explained by the cookie acceptance behaviour of the "I don't care about cookies" extension. Interestingly, "www.sky.at", which is taking second place in fig. 4.4a and fig. 4.4b, is only in 12th place based on the data gathered by this script. It seems the "I don't care about cookies" extension entirely blocks Sourcepoint's CMP, which is used on Sky's website. Therefore, all cookies associated with Sourcepoint are missing from data gathered in this particular script execution.

Further comparison

URLs:

There are a few URLs, which were only found with a specific script, while others were found regardless of a script's execution with or without extension.

A single website appeared only in the data gathered without extension: "www.ubm-development.com". The cookie gathered from this site indicated a redirect based on geolocation. There is no conclusive explanation, why this cookie was found on this particular script run. To verify if it has anything to do with the script's variable, the web scraper without extensions was executed on this website alone for a few more times. However, the cookie was not set by the website again.

The data gathered from the script with Consent-O-Matic extension contained 12 URLs, which were neither included in the data from the script without extensions nor in the data from the execution with "I don't care about cookies" extension. The "I don't care about cookies" extension run yielded 28 such URLs. These could point to differences in the cookie acceptance behaviour of the extensions.

As can be seen in fig. 4.4, the number of gathered cookies increases overall with the use of extensions. Most changes in the number of cookies between those three figures can be explained by differences in cookie acceptance behaviour of the used extension. The website with the highest number of cookies never changes, but scripts with extensions still gathered more cookies while scraping it than the script without extensions did.

Cookies:

There are 13 unique cookies, which were only present on script execution without extensions. One of those is the aforementioned geolocation cookie by "www.ubm-

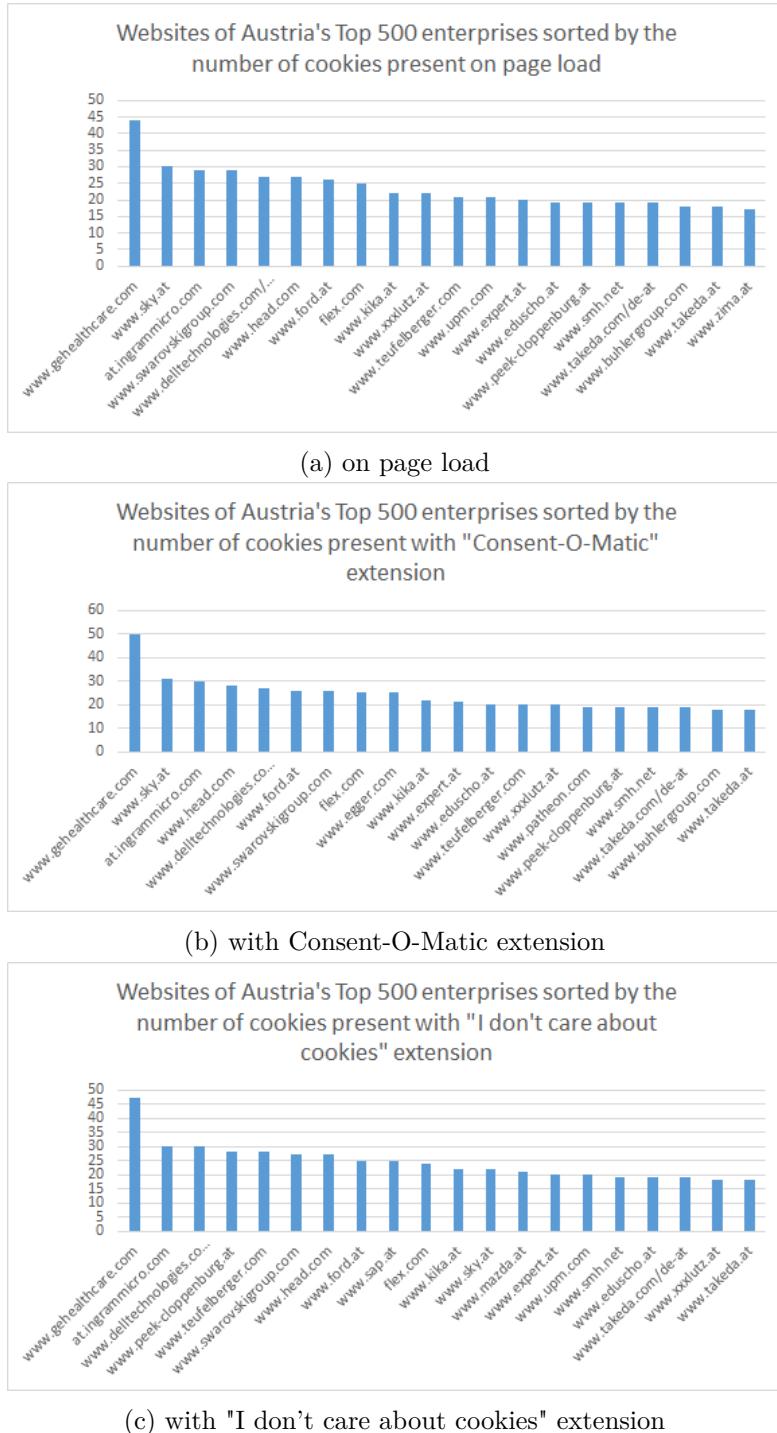


Figure 4.4: Comparison of first 20 URLs sorted descending by the number of cookies present with and without extensions

4. WEB SCRAPING: UTILISATION OF USER TRACKING TECHNOLOGIES BY TOP 500 AUSTRIAN ENTERPRISES

development.com". Others are cookies by the CMP "consentmanager"³⁷ and real-time personalisation service "TRBO",³⁸ both of them include a generated ID and are therefore unique for each script execution. This also explains most of the 107 cookies, which were found exclusively in the script run featuring Consent-O-Matic. Another reason is the suspected change of cookies set by HotJar, as there are now cookies named "_hjSession_[x]" and "_hjSessionUser_[x]", where [x] stands for a generated number, which have not been present in any of the earlier gathered data sets. The script run with "I don't care about cookies" extension yielded 97 cookies not found with the other two scripts. Most of them were, again, the same cookies but with other generated numbers attached. One page, which must have received cookie consent from the "I don't care about cookies" extension, added tracking cookies associated with the analytics tool "Lucky Orange".³⁹

4.3.2 JavaScript files

As the web scraper for script files scraped first- as well as third-party scripts, analysis was a little more difficult. JS-files are especially interesting to detect fingerprinting activity. As it is rather unusual that fingerprinting is done by first parties, most relevant information for chapter 5 was extracted from section 4.3.3. CNAME tracking which is only done via first-party domains is unfortunately not detected by the type of scripts used. Nonetheless, the following statistics are based on the gathered data of the self-implemented scripts and are used complementary to the data gathered by webXray.

The data gathered by the web scraping files was once again filtered for duplicates (identical URL and name of the script source) and rows containing only error messages. In fig. 4.5 the 20 URLs with the highest number of first- and third-party script sources present are compared. The number of JS-files detected by the web scraper increased with the use of extensions. Instead of a total of 10,206 files on 567 unique URLs, the script found 10,553 files with Consent-O-Matic and 10,665 files with "I don't care about cookies" extension. The number of pages stayed the same. Of the files found 8,859, 9,156 and 9,170, respectively, were unique. Some files (158 without extension, 199 with Consent-O-Matic, 200 with "I don't care about cookies") could not be considered while counting unique script sources as the name of the script source exceeded 255 characters and therefore the limit of the COUNTIF function of Microsoft Excel. As this data is only used to complement the findings in section 4.3.3 no effort to overcome this limitation was made.

4.3.3 webXray reports

webXray was able to scrape 564 websites; 9 URLs were removed before a scraping attempt was made due to issues with the URL itself; 43 scraping attempts failed.

³⁷<https://help.consentmanager.net/books/cmp/page/cookies-set-by-the-cmp>, last visited 2021-11-23

³⁸<https://www.trbo.com/en/>, last visited 2021-11-23

³⁹<https://help.luckyorange.com/article/173-what-cookies-does-lucky-orange-set>, last visited 2021-11-23

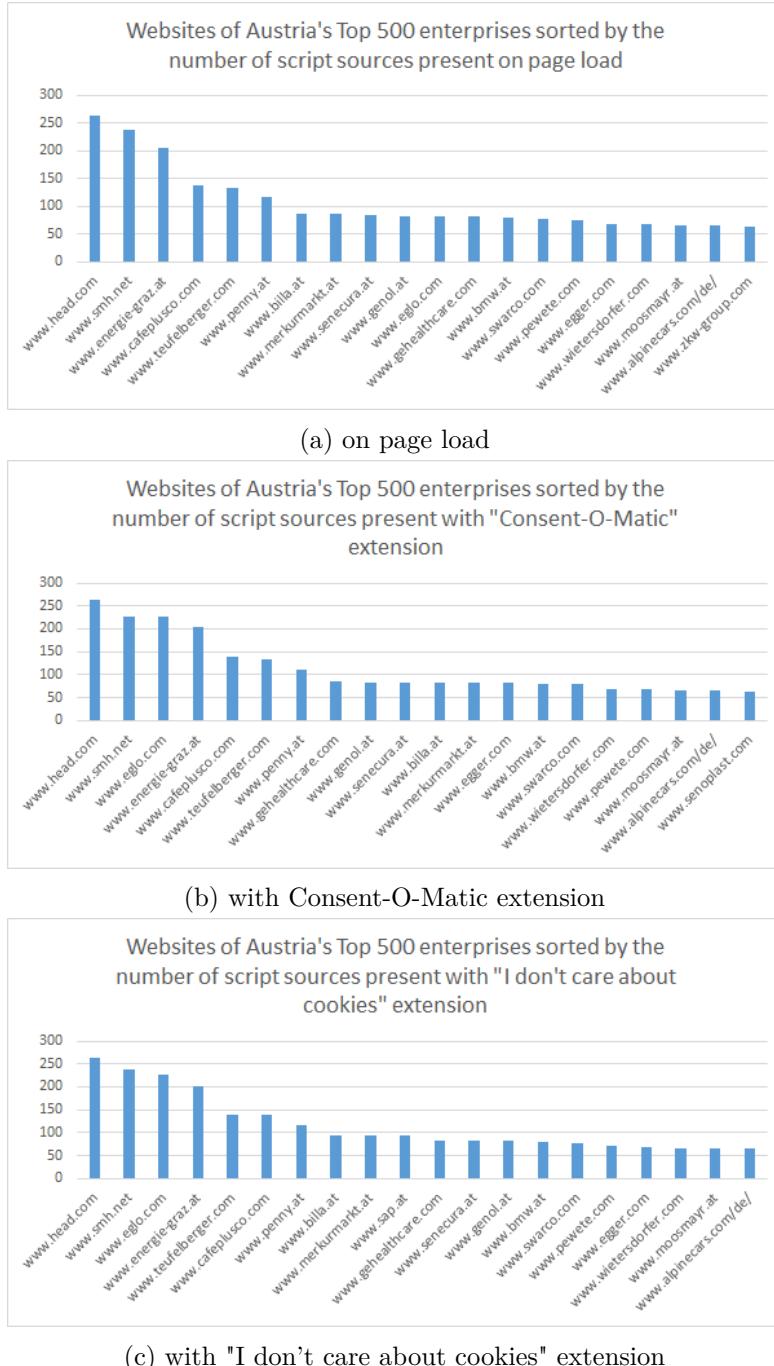


Figure 4.5: Comparison of first 20 URLs sorted descending by the number of script sources present with and without extensions

4. WEB SCRAPING: UTILISATION OF USER TRACKING TECHNOLOGIES BY TOP 500 AUSTRIAN ENTERPRISES

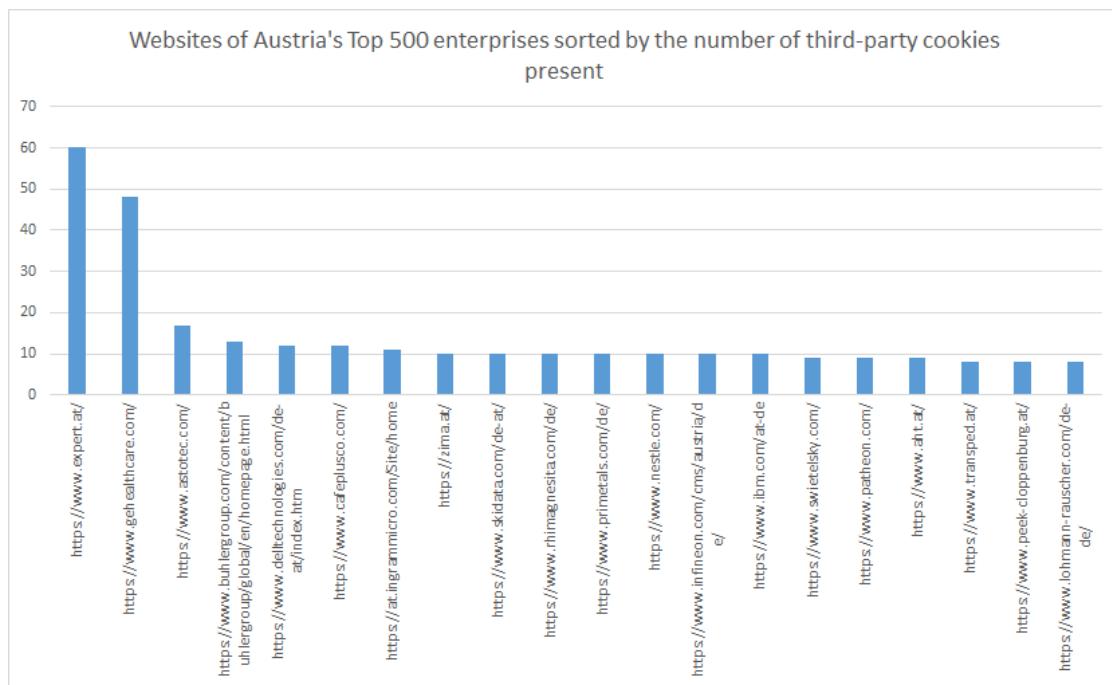


Figure 4.6: First 20 URLs sorted descending by the number of third-party cookies sources found by webXray

Third-party cookies

On these websites 2824 cookies were found, 631 of them being third-party cookies. That first number closely matches the number of first-party cookies found in section 4.3.1. The URL with the most third-party cookies present was "www.expert.at", followed by "www.gehealthcare.com" and "www.astotec.com". The first two of them have already been in the Top 20 URLs regarding first-party cookies. Other than that, the Top 20 URLs with respect to third-party cookie presence have very little similarity to the 20 URLs with the highest first-party cookie count. All 20 websites are shown in fig. 4.6.

The domain responsible for the most third-party cookies is ".linkedin.com" with 162 cookies, which is nearly double the amount of third-party cookies found on the domain coming in second place ("youtube.com"). There are only 10 third-party domains overall, which place 10 or more cookies.

Third-party scripts

One of the automatically generated reports created by webXray focuses on third-party scripts. For the 564 successfully scraped websites, the report includes 500 unique third-party scripts by 124 unique domains. Figure 4.7 shows the Top 15 third-party scripts, all of them were found on at least 4% of all successfully scraped websites. While "https://www.googletagmanager.com/gtm.js" was the script found most, "netdna-ssl.com"

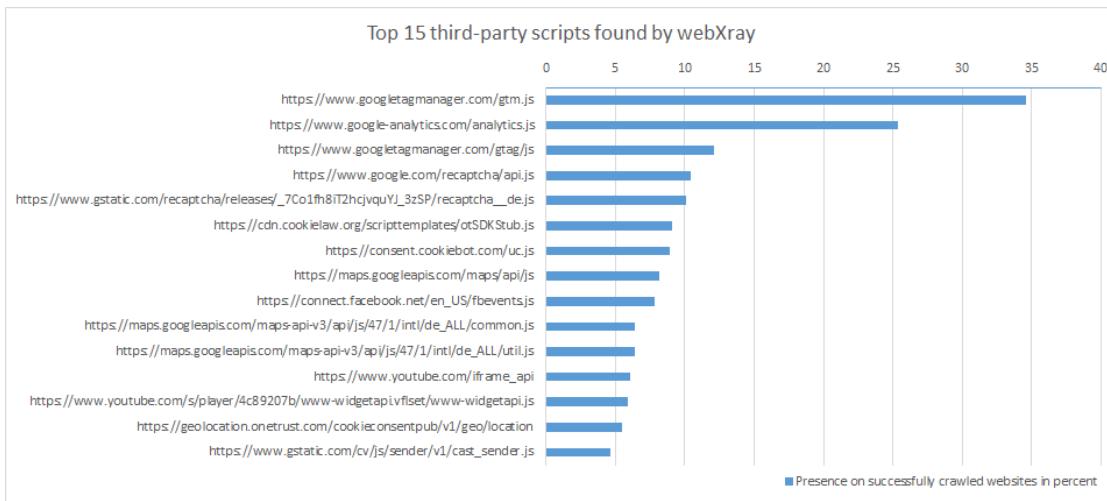


Figure 4.7: Top 15 third-party scripts found by webXray, sorted by percentage of presence on websites provided by Austria's Top 500 enterprises

contributed the highest number of unique JS-files to the database. The first is followed by "https://www.google-analytics.com/analytics.js", which was found on 25.35% of all successfully scraped websites, and "https://www.googletagmanager.com/gtag/js" with a presence on 12.06% of those pages. The latter is followed by 53 unique scripts contributed by "googleapis.com" and 32 unique scripts associated with "facebook.com". webXray matches certain domains to their owner, meaning services which provide those scripts, e.g. "netdna-ssl.com" to StackPath. The gathered data contained 43 known domain owners. webXray even offers insight on "owner lineage", information on the company to which a service is associated and if the company itself is owned by another company, e.g. Google Analytics, belonging to Google and Google being part of the Alphabet group.

Fingerprinting:

A full analysis on utilised fingerprinting methods is not feasible within the scope of this thesis. However, based on the information about third-party scripts gathered by webXray, a comparison of found domains to known fingerprinting domains mentioned in existing research (see also section 2.3.1) was done. Their presence alone does not conclusively prove the actual application of fingerprinting techniques, but they give a good indication, if fingerprinting is even possible to Austria's Top 500 enterprises.

Google Analytics, which Al-Fannah et al. named as the "*most widely used fingerprinting third-party*"⁴⁰ in 2018, was present with 5 different scripts on websites of Austria's Top 500 enterprises. Of the other Top 10 third-party domains named by Al-Fannah, 4 can also be found in the data gathered by webXray: doubleclick.net, google.com, quantserve.com and bing.com. Acar et al. identified 13 font fingerprinting scripts in 2013 and explicitly named them in their paper.⁴¹ A comparison of script names showed that only one of

⁴⁰[Al-Fannah et al., 2018], p. 489

⁴¹[Acar et al., 2013], p. 1135

4. WEB SCRAPING: UTILISATION OF USER TRACKING TECHNOLOGIES BY TOP 500 AUSTRIAN ENTERPRISES

them is present on websites provided by Austria's Top 500 enterprises. The script named "cc.js" was found 4 times, none of them being connected to the provider named by Acar et al. It is still possible that those scripts have been renamed or other font fingerprinting scripts are present, however, this is not verifiable within the scope of this work. A year later, Acar et al. looked at canvas fingerprinting and yet again named a number of canvas fingerprinting domains.⁴² None of those scripts could be found in the data gathered by webXray. Again, it is still possible that those scripts have been renamed or other canvas fingerprinting scripts are present. Especially, because "addthis.com", the provider named by Acar et al. as number one provider of canvas fingerprinting in their paper⁴³, is present with other scripts.

DOM storage

WebXray also inspects the browser's Document Object Model (DOM) storage with its two main storage types: localStorage and sessionStorage. HTML5 localStorage tracking methods have been discussed in section 2.3.1. On the 564 successfully scraped websites webXray found a total of 2421 objects in DOM storage. Of those 1660 were stored in the HTML5 localStorage on 336 websites; 531 were placed by third parties. This means there were localStorage entries on 59.57% of all websites successfully scraped by webXray.

As can be seen in fig. 4.9, the website with the most DOM storage entries, is by far "<https://www.eaton.com/at/de-de.html>". Followed by the websites of Buhler Group and Head. Most of "<https://www.eaton.com/at/de-de.html>"'s DOM storage entries are obfuscated and cannot be matched to any service or tool. They also include entries belonging to "addthis.com" and Adobe. As there is little information on HTML5 localStorage entries, I could not verify what these entries are actually used for. Figure 4.10 shows the Top 10 DOM storage entries, all found on at least 25 individual websites of Austria's Top 500 enterprises. The number one entry is "modernizr", which can be explained by the nature of this tool. Modernizr is used to detect supported features in users' browsers, HTML5 localStorage being one of them.⁴⁴ It allows companies to deliver websites that work with a given browser's feature set, even if it does not support certain technologies.

The second entry is set by "<https://www.youtube.com>" or "<https://www.youtube-nocookie.com>". YouTube states that the use of their "nocookie" domain, when embedding videos on a website, leads to no visitor information being stored by YouTube unless the video itself is played (see fig. 4.8). Based on the findings of this thesis, and backed up by findings of heise.de,⁴⁵ YouTube's claim is false as data is stored in HTML5 localStorage without a video being played. This also includes other entries in fig. 4.10, which are all starting with "yt".

⁴²[Acar et al., 2014], p. 679

⁴³[Acar et al., 2014], p. 678

⁴⁴<https://modernizr.com/docs/>, last visited 2021-11-27

⁴⁵<https://www.heise.de/select/ct/2016/1/1451711441689162>, last visited 2021-11-27

"req" is placed by Cookiebot, its functionality is unclear. Google Recaptcha's "rc::a" takes fourth place; it is used "to read and filter requests by bots" according to Cookiedatabase.org.⁴⁶ Adobe's "com.adobe.reacto.dataElementCookiesMigrated" is used to migrate cookies between the outdated Adobe Dynamic Tag Management and Adobe Launch. This information could not be verified on Adobe's website, but was found in several cookie lists appended to cookie policies, including the one by Danske Bank.⁴⁷ The tenth entry is set by several first parties, which have no visible similarities. Its usage is therefore unknown.

HTTP requests and responses

A total of 47,676 HTTP requests, including 14,401 requests by third parties, were observed by webXray. About 99.4% of those requests received a response. Libert categorised requests in seven element types based on file extensions: JavaScript, PHP scripts, CGI scripts, images, fonts and CSS, JSON.⁴⁸ JavaScript has already been discussed in detail in section 4.3.2 and analysing PHP and CGI scripts as well as requested fonts and CSS-files lies outside the scope of this work. Data collected by webXray shows that none of the analysed websites requests additional JSON content on page load.

In total, Austria's Top 500 enterprises request 20,069 image files, 5,111 of those are requested from third parties. Libert provides a list of known tracking images⁴⁹, of these only one was found on Austria's Top 500 enterprise's websites too. The Google Analytics tracking pixel "_utm.gif", which was the most requested image in Libert's research, was found on seven different websites in the data gathered using webXray. The same file is also used by other Google services. Libert names DoubleClick⁴⁹, however, all DoubleClick tools have been renamed by the time this work was written.⁵⁰ The Google Analytics tracking pixel is used to transfer all gathered data from a client to the server via a simple HTTP GET request.⁵¹

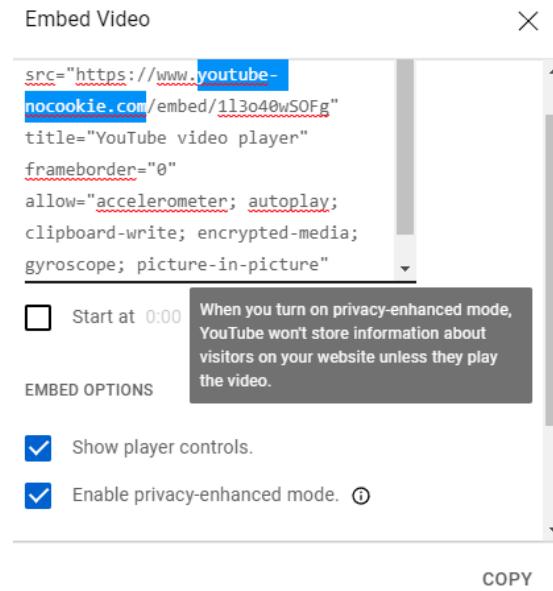


Figure 4.8: YouTube's dialog for video embed code generation

⁴⁶<https://cookiedatabase.org/cookie/google-recaptcha/rca/>, last visited 2021-11-27

⁴⁷<https://danskebank.com/-/media/pdf/cookies/cookie-list-16062020.pdf>, last visited 2021-11-27

⁴⁸[Libert, 2015], p. 3552

⁴⁹[Libert, 2015], p. 3549

⁵⁰<https://www.reuters.com/article/us-alphabet-google-advertising/google-retires-doubleclick-adwords-brand-names-idUSKBN1JN0EH>, last visited 2021-11-27

⁵¹<https://developers.google.com/analytics/resources/concepts/gaConceptsTrackingOverview>, last visited 2021-12-02

4. WEB SCRAPING: UTILISATION OF USER TRACKING TECHNOLOGIES BY TOP 500 AUSTRIAN ENTERPRISES

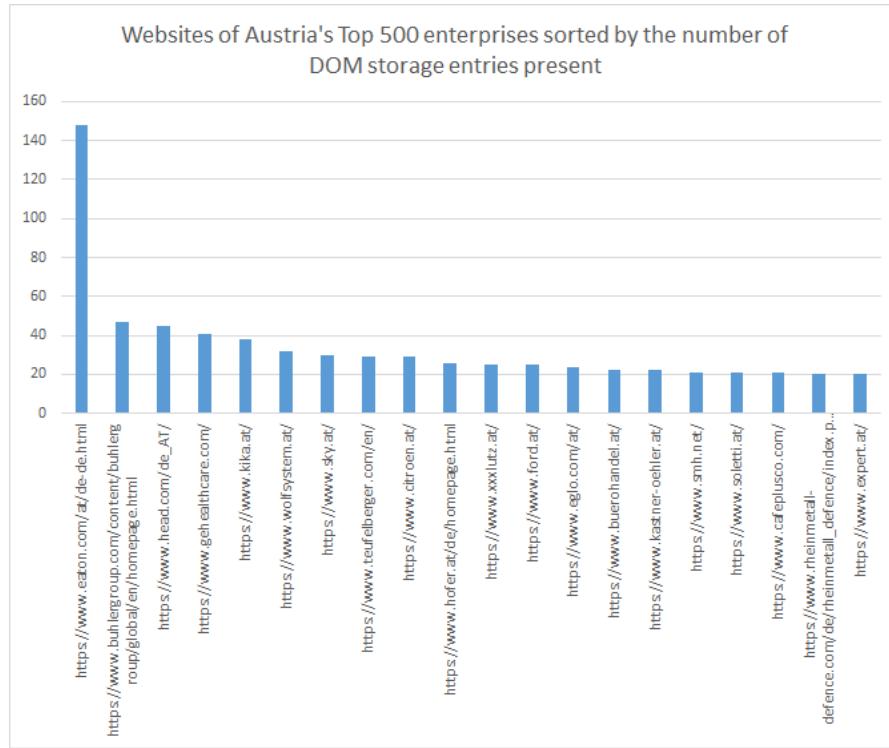


Figure 4.9: First 20 URLs sorted descending by the number of DOM storage entries found by webXray

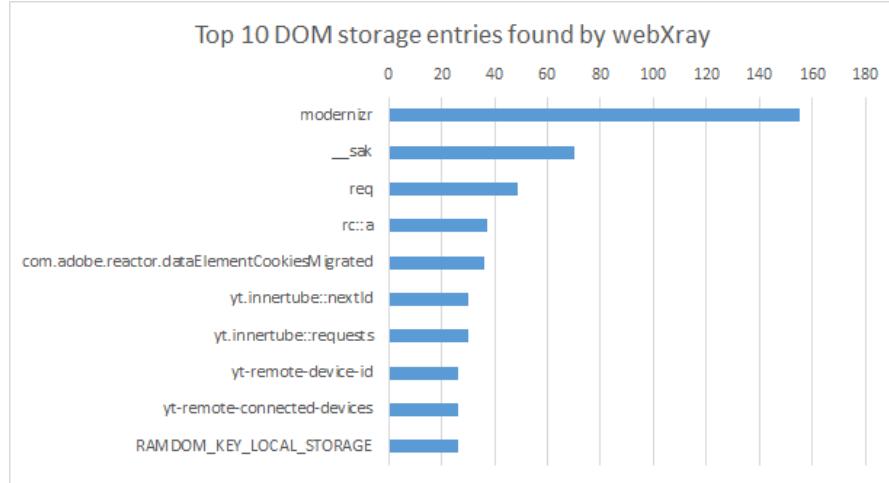


Figure 4.10: Top 10 third-party scripts found by webXray, sorted by their presence in absolute numbers on websites provided by Austria's Top 500 enterprises

CHAPTER 5

Legal analysis: User tracking in Austria and the European Union

5.1 Methodology

The following chapter is intended to answer the following questions:

3rd research question:

In which ways are these technologies covered by Austrian or European law? Which technologies in use pose potential threats to users' privacy and are currently not regulated?

The starting point for the legal analysis was a book called "E-Commerce- und Internetrecht" by Straube and Fina.¹ Its last volume was published in 2010, however, it still provides good guidance. Straube and Fina gathered legislative texts regarding e-commerce and/or the Internet from the EU and Austria. The fifth volume was published following an initiative of the Stanford-Vienna Transatlantic Technology Law Forum. While part of the legislation named in the book is still in force, like the Directive on privacy and electronic communications,² others have (partly) been replaced or are no longer in force, like Directive 2006/24/EC (commonly known as "Data Retention Directive").³

Based on the legal texts mentioned in "E-Commerce- und Internetrecht", I checked for the current status of said texts and their relevance to the results of chapters 3 and 4. Whenever a text I considered relevant was still in force, it was added to the list in section 5.2. If it was repealed or replaced, I started looking for newer resources covering the same area. After finishing the initial analysis, I searched the respective

¹[Straube and Fina, 2010]

²[Directive on privacy and electronic communications, 2002]

³[Directive 2006/24/EC, 2006]

texts for references to other legislation. Finally, I looked at recent court rulings citing the found legal texts, especially in regard to the findings of chapters 3 and 4. EU legislation was researched on the EUR-lex online service provided by the Publications Office of the European Union⁴; European case law was looked up using CURIA, the website of the Court of Justice of the European Union (CJEU);⁵ all Austrian legal documents can be found in the legal information system of the Republic of Austria (Rechtsinformationssystem des Bundes (RIS)).⁶ The results of this first analysis can be found in the following sections. Section 5.3 is sorted based on tracking methods and the order they were named in section 2.3; with the exception of third-party cookies which are discussed together with other cookie-related topics in section 5.3.1.

5.2 Overview of relevant legal texts

All legislation below is sorted by the year it first came into effect. Court rulings are sorted by the year their judgment was issued.

5.2.1 European Union

EU legislation

*"All European Union law is derived from the treaties between the EU Member States. The treaties function as the EU's de facto constitution, defining both the allocation of powers between the EU and the Member States and the allocation of powers among the EU's institutions."*⁷ One of those treaties is the Treaty on the Functioning of the European Union. It states: *"To exercise the Union's competences, the institutions shall adopt regulations, directives, decisions, recommendations and opinions."*⁸ It also regulates which of the above mentioned texts are legally binding and/or directly applicable and which are not. While directives, which are addressed to the member states, must be transferred into national legislation, regulations and decisions are legally binding and directly applicable. The difference between regulations and decisions is that decisions are only legally binding to those to whom they are addressed while regulations must be applied across the European Union. Recommendations and opinions are not binding, they are mere instruments of guidance.

Directive on electronic commerce⁹:

Since it came into force in July 2000, the Directive on electronic commerce attempts to *"ensure legal certainty and consumer confidence"*.¹⁰ To achieve this the *"Directive must lay down a clear and general framework to cover certain legal aspects of electronic commerce"*

⁴<https://eur-lex.europa.eu/content/welcome/about.html>, last visited 2021-12-02

⁵https://curia.europa.eu/jcms/jcms/_j_6/en/, last visited 2021-01-02

⁶<https://www.ris.bka.gv.at/default.aspx>, last visited 2021-12-02

⁷<https://libguides.law.illinois.edu/EU/treaties>, last visited 2021-11-28

⁸[Treaty on the Functioning of the European Union, 2012], Art. 288

⁹[Directive on electronic commerce, 2000]

¹⁰[Directive on electronic commerce, 2000], Rec. 7

in the internal market."¹⁰ As its Article 6 and 7 regulate commercial communication, they are especially relevant to section 5.3.4 covering the legal situation of newsletter distribution.

Charter of Fundamental Rights of the European Union (CFR)¹¹:

Created in October 2000 by the European Convention and ratified in December of the same year, the Charter of Fundamental Rights of the European Union became legally binding in December 2009. Between 2000 and 2002 there was some uncertainty about the legal status of the CFR.¹² Since the Treaty of Lisbon came into effect, it has the same legal value as European treaties.¹³ For the purpose of thesis, Article 7 and 8 are especially interesting and have been cited in several relevant legal texts and court cases, for example in "Maximillian Schrems v Data Protection Commissioner".¹⁴

Directive on privacy and electronic communications¹⁵:

Directive 2002/58/EC, with its official nickname "Directive on privacy and electronic communications", is sometimes also known as "E-Privacy Directive"¹⁶, which can be misleading as Directive 2009/136/EC is also known as such. The Directive on privacy and electronic communications was meant to harmonise EU citizens' right to privacy, *"with respect to the processing of personal data in the electronic communication sector and to ensure the free movement of such data and of electronic communication equipment and services in the Community."*¹⁷ When the directive came into effect, it complemented Directive 95/46/EC,¹⁸ which has since been repealed by the GDPR.^{17,19} Article 5(3) is of special interest to this work as it regulates the use of browser cookies and other storage-based user tracking methods.

E-privacy Directive²⁰:

Directive 2009/136/EC, commonly referred to as "E-privacy Directive"²¹, came into effect on 26 November 2009. It amended the Directive on privacy and electronic communications, the Universal Service Directive,²² which has since been repealed by Directive (EU) 2018/1972²³, and the Regulation on consumer protection cooperation,²⁴ which has since been repealed by Regulation (EU) 2017/2394²⁵. Relevant changes to aforementioned

¹¹[Charter of Fundamental Rights of the European Union, 2012]

¹²[Menéndez, 2002], p. 473 ff.

¹³[Treaty of Lisbon, 2007], Art. 6

¹⁴[Schrems (C-362/14), 2015], Sect. 1

¹⁵[Directive on privacy and electronic communications, 2002]

¹⁶[Debusseré, 2005], p. 80

¹⁷[Directive on privacy and electronic communications, 2002], Art. 2

¹⁸[Directive 95/46/EC, 1995]

¹⁹[General Data Protection Regulation, 2016], Rec. 171

²⁰[Directive 2009/136/EC, 2009]

²¹https://edps.europa.eu/data-protection/data-protection/glossary/e_en#e-privacy_directive2009-136-ec, last visited 2021-11-28

²²[Universal Service Directive, 2002]

²³[Directive (EU) 2018/1972, 2018]

²⁴[the Regulation on consumer protection cooperation, 2004]

²⁵[Regulation (EU) 2017/2394, 2017]

5. LEGAL ANALYSIS: USER TRACKING IN AUSTRIA AND THE EUROPEAN UNION

Article 5(3) of the Directive on privacy and electronic communications were made with this directive²⁶.

General Data Protection Regulation (GDPR)²⁷:

As mentioned at the beginning of this section, regulations are one of the instruments available to the EU to create texts that are legally binding as well as directly applicable across the Member States. Before the General Data Protection Regulation came into effect, the subject of data protection was governed by Directive 95/46/EC. The "*fragmentation in the implementation of data protection across the Union*"²⁸ caused by said directive and the harmonisation of those fragmented implementations were reasons for creating the GDPR.²⁹ The scope of the GDPR is limited to personal data of natural persons.³⁰ What is considered personal data is defined in Article 4(1) of the GDPR. Some parts of the GDPR mandate a specification by the Member states, others allow for additional rules.³¹ Therefore, Member States still have national data protection acts. The Austrian one is mentioned in section 5.2.2.

EU case law

The Court of Justice of the European Union "*constitutes the judicial authority of the European Union*"³² and is divided into 2 courts: the European Court of Justice (ECJ) and the General Court. The ECJ is constituted by one judge from each member state, with an additional 11 advocates general. The General Court consists of two judges from every EU country.³³ The ECJ can sit as full court, or in one of two constellations: in Grand Chamber, consisting of 15 judges, or in Chambers of three to five judges. Which form is chosen depends on the complexity and importance of the case.³⁴ While the ECJ deals mainly with member states or their respective courts, the General Court "*rules on actions for annulment brought by individuals, companies and, in some cases, EU governments.*"³³

Maximillian Schrems v Data Protection Commissioner³⁵:

Maximilian Schrems challenged the refusal of the Irish Data Protection Commissioner (DPC) to investigate his complaint. His initial complaint was made based on revelations by Edward Snowden suggesting that personal data of EU citizens stored in the U.S. could be accessed by U.S. intelligence authorities. He asked that data transfer between

²⁶[Directive 2009/136/EC, 2009], Art. 2

²⁷[General Data Protection Regulation, 2016]

²⁸[General Data Protection Regulation, 2016], Rec. 9

²⁹[General Data Protection Regulation, 2016], Art. 10 & 11

³⁰[General Data Protection Regulation, 2016], Art. 1(1)

³¹[General Data Protection Regulation, 2016], Rec. 8

³²https://curia.europa.eu/jcms/jcms/J_02_6999/en/, last visited 2021-11-30

³³https://europa.eu/european-union/eu-institutions-law-budget/institutions-and-bodies/institutions-and-bodies-profiles/court-justice-european-union-cjeu_en, last visited 2021-11-30

³⁴https://curia.europa.eu/jcms/jcms/J_02_7024/en/, last visited 2021-11-30

³⁵[Schrems (C-362/14), 2015]

Facebook Ireland and Facebook Inc. should be ceased due to the violation of his right to data protection.³⁶ The Irish High Court referred his case to the CJEU in 2014, judgment was passed by the Grand Chamber of the ECJ in 2015. The case is known as "Maximillian Schrems v Data Protection Commissioner", "Schrems" or sometimes "Schrems I". The preliminary ruling of the ECJ invalidated the EU Commission's decision of legal adequacy of the U.S.-EU Safe Harbor Framework³⁷ and led to the creation of the EU-U.S. and Swiss-U.S. Privacy Shield Frameworks.³⁸

Patrick Breyer v Bundesrepublik Deutschland³⁹:

In 2014, the German Federal Court referred two questions to the CJEU. The first question regarded Article 2(a) of Directive 95/46/EC and its interpretation of dynamic IP addresses as personal data. The second question dealt with specifics on the interpretation of Article 7(f) of Directive 95/46/EC.⁴⁰ The case, known as "Patrick Breyer v Bundesrepublik Deutschland" or "Breyer", is considered special, because *"for the first time the CJEU focused on the subjective aspect of identification by a concrete controller or processor of personal data with regard to the means available to them for identifying a given personal data subject."*⁴¹ The ruling, together with other court cases (e.g. Scarlet Extended⁴²), have most likely led to the explicit mention of IP addresses as a form of data which, in some cases, is considered identifying in Recital 30 of the GDPR.

Fashion ID GmbH & Co.KG v Verbraucherzentrale NRW eV.⁴³:

The Higher Regional Court Düsseldorf (Oberlandesgericht Düsseldorf) referred 6 questions to the CJEU for a preliminary ruling. The ECJ's Second Chamber issued its judgement in July 2019. The questions refer to Directive 95/46/EC, the predecessor of the GDPR, which was still in force by the time the case was opened. The court decided that, based on Directive 95/46/EC, national law can allow *"consumer-protection associations to bring or defend legal proceedings against a person allegedly responsible for an infringement of the protection of personal data."*⁴⁴ This decision is also in accordance with Article 80 of the GDPR. Other than that the questions are focused on the legal implications of embedded third-party code, in particular Facebook's "Like"-Button, and the responsibilities of first and third parties regarding informing and requesting consent from users about the collection of personal data.

³⁶<https://iapp.org/resources/article/schrems-i/>, last visited 2021-11-30

³⁷[Schrems (C-362/14), 2015], Para. 106

³⁸<https://www.ftc.gov/tips-advice/business-center/privacy-and-security/u.s.-eu-safe-harbor-framework>, last visited 2021-11-30

³⁹[Judgment - Breyer (C-582/14), 2016]

⁴⁰[Application - Breyer (C-582/14), 2015]

⁴¹<https://www.lexology.com/library/detail.aspx?g=34d7c88e-a350-497c-93ce-adef3fb6484f>, last visited 2021-11-30

⁴²[Scarlet Extended (C-311/18), 2011]

⁴³[Fashion ID (C-40/17), 2019]

⁴⁴[Fashion ID (C-40/17), 2019], Para. 63

Bundesverband der Verbraucherzentralen und Verbraucherverbände - Verbraucherzentrale Bundesverband e.V. v Planet49 GmbH.⁴⁵:

The German Federal court referred a list of questions to the CJEU in this case, commonly called "Planet49". The outcome significantly clarified the EU's position on cookie consent. CMPs often used pre-ticked checkboxes when asking for consent for all types of cookies, but Recital 32 of the GDPR states clearly that pre-ticked checkboxes do not fulfill the requirements for active consent. The judgement in this case takes that into consideration.⁴⁶ It also specifically clarifies that consent is needed for storing any kind of information or accessing any kind of already stored information on a user's terminal equipment, based on Article 5(3) of the Directive on privacy and electronic communications.⁴⁷ The German Federal Court asked explicitly if this article had to be interpreted differently for personal data. The last question regarded the amount of information which must be made available about cookies.⁴⁸ The ECJ stated: "*[...] the information that the service provider must give to a website user includes the duration of the operation of cookies and whether or not third parties may have access to those cookies.*"⁴⁹

Data Protection Commissioner v Facebook Ireland Ltd, Maximillian Schrems⁵⁰:

In 2015, the Irish DPC informed Maximillian Schrems that Facebook relied on standard contractual clauses (SCC) (see section 5.3.7 for more information) for data transfer and not on the now invalidated U.S.-EU Safe Harbor Framework, thereby, rendering the ruling of the ECJ in "Schrems I" irrelevant. Maximillian Schrems addressed a reformulated complaint to the DPC shortly after, which led to a lawsuit by the DPC against Facebook Ireland and Maximillian Schrems. The DPC had to file such a suit to comply to the legal process defined in Paragraph 65 of "Schrems I" for situations "*where the national supervisory authority considers that the objections advanced by the person who has lodged with it a claim concerning the protection of his rights and freedoms in regard to the processing of his personal data are well founded*". This case was heard by the Irish High Court, which then referred a total of 11 questions to the CJEU.⁵¹ The judgment issued by the ECJ on this case invalidated the EU-U.S. Privacy Shield framework.⁵² The case is known as "Data Protection Commissioner v Facebook Ireland Ltd, Maximillian Schrems", "Facebook Ireland and Schrems", or simply "Schrems II".⁵³

⁴⁵[Planet49 (C-673/17), 2019]

⁴⁶[Planet49 (C-673/17), 2019], Para. 62

⁴⁷[Planet49 (C-673/17), 2019], Para. 66-71

⁴⁸[Planet49 (C-673/17), 2019], Para. 37

⁴⁹[Planet49 (C-673/17), 2019], Para. 81

⁵⁰[Facebook Ireland and Schrems (C-311/18), 2020]

⁵¹<https://noyb.eu/en/project/eu-us-transfers>, last visited 2021-11-30

⁵²[Facebook Ireland and Schrems (C-311/18), 2020], Para. 201

⁵³<https://www.bfdi.bund.de/DE/Fachthemen/Inhalte/Europa-Internationales/Auswirkungen-Schrems-II-Urteil.html>, last visited 2021-12-04

5.2.2 Austria

Austrian legislation

Allgemeines Bürgerliches Gesetzbuch (ABGB)⁵⁴:

The General Austrian Civil Code (Allgemeines Bürgerliches Gesetzbuch) came into effect when Austria was still the Austrian Empire. Its first draft was written by a commission led by Karl Anton von Martini; it was revised by a commission led by Franz von Zeiller and sanctioned by Emperor Francis I in 1811.⁵⁵ Over time, there have been several amendments and reforms. The Austrian Data Protection Authority (Datenschutzbehörde) names two relevant sections of the General Austrian Civil Code with respect to data protection, namely Section 16 and Section 1328a.⁵⁶ They can be viewed as legal groundwork for any kind of data protection legislation in Austria, similarly to Article 7 and 8 of the CFR in the EU.

Datenschutzgesetz (DSG)⁵⁷:

Until May 25 2018, the Austrian Data Protection Act (Datenschutzgesetz) was known as "Datenschutzgesetz 2000 (DSG 2000)". It was the national transposition of EU Directive 95/46/EC. When the GDPR came into effect, it rendered most of Austria's DSG 2000 obsolete.⁵⁸ The current Datenschutzgesetz complements the GDPR; it provides concrete information on the applicability of the GDPR in Austria and names the supervisory authority with which GDPR complaints can be filed.⁵⁹ Article 2 of the DSG also includes additional rules where the GDPR left room for national legislation to do so.

E-Commerce Gesetz (ECG)⁶⁰:

The Austrian E-Commerce Act (E-Commerce Gesetz) transposes the Directive on electronic commerce of the EU⁶¹ into national law. There exists an official English translation of this act in the legal information system of the Republic of Austria (RIS).⁶²

Telekommunikationsgesetz 2021 (TKG21)⁶³:

On 1st November 2021 the new Austrian Telecommunications Act TKG21 came into effect. It, like its predecessor from 2003, includes the national transposition of EU Directive 2002/58/EC. Especially the implementation of Article 5(3) of the Directive on privacy and electronic communications in Section 165 Paragraph 3 is cause of ongoing discussion. The Viennese law office "Geistwert" published an article questioning this

⁵⁴[Allgemeines bürgerliches Gesetzbuch, JGS 946/1811 idF I 175/2021, 1811]

⁵⁵[Meissel, 2017], p. 112

⁵⁶<https://www.dsbs.gv.at/recht-entscheidungen/gesetze-in-oesterreich.html>, last visited 2021-11-28

⁵⁷[Datenschutzgesetz, BGBl. I 165/1999 idF I 14/2019, 1999]

⁵⁸[Datenschutzgesetz 2000, BGBl. I 165/1999 idF I 120/2017, 1999], Sect. 70

⁵⁹[Datenschutzgesetz, BGBl. I 165/1999 idF I 14/2019, 1999], Art. 2

⁶⁰[E-Commerce-Gesetz, BGBl. I 152/2001 idF I 148/2020, 2001]

⁶¹[Directive on electronic commerce, 2000]

⁶²https://www.ris.bka.gv.at/Dokument.wxe?Abfrage=Erv&Dokumentnummer=ERV_2001_1_152, last visited 2021-12-01

⁶³[Telekommunikationsgesetz 2021, BGBl. I Nr. 190/2021, 2021]

section's restriction to personal data and stated that this could be considered a breach of EU law.⁶⁴

Austrian court cases and DSB decisions

There are several Austrian court cases regarding GDPR complaints, where important questions were referred to the CJEU for preliminary ruling in the last year. Judgment on those cases is still outstanding, but they could potentially lead to further clarification of ambiguous worded parts of the GDPR. One of those cases originated, once again, from a complaint by Maximilian Schrems against Facebook Ireland Ltd.⁶⁵

The Austrian Data Protection Authority (Datenschutzbehörde (DSB)) is Austria's national supervisory authority responsible for monitoring the application of the GDPR.⁶⁶ In this role the DSB decided in January 2022 that the continuous use of Google Analytics after the ECJ judgement on "Schrems II" violates the GDPR.⁶⁷ Similar decisions have been reached shortly before by the European Data Protection Supervisor⁶⁸ and shortly after by the French Data Protection Authority.⁶⁹

5.3 Applicability & Interpretation

The sections below are based on the findings of chapters 3 and 4. Some areas could not be investigated further as the information gathered on certain topics was insufficient for a legal analysis. E.g. the survey's questions about third-party advertisements: No tools were named by any of the respondents and chapter 4 did not cover that area at all. Therefore, the gathered material was insufficient to cover utilised user tracking methods in this regard. There is no specific section on online shops because the tools mentioned in the survey either use cookies⁷⁰ (which are covered below), do not disclose their tracking mechanisms or do not collect personal data of their client's users at all.⁷¹

Before going further into detail, I want to cite Articles 7 and 8 of the Charter of Fundamental Rights of the European Union, which were previously mentioned, as they can be considered fundamental to all subsequent arguments.

⁶⁴<https://geistwert.at/neue-oesterreichische-cookie-bestimmung-im-tkg-2021-weiterhin-europarechtswidrig/>, last visited 2021-12-01

⁶⁵[Maximilian Schrems v Facebook Ireland Ltd (C-446/21), 2021]

⁶⁶[Datenschutzgesetz, BGBl. I 165/1999 idF I 14/2019, 1999], Sect. 31

⁶⁷<https://www.dsbs.gv.at/dam/jcr:c1eb937b-7527-450c-8771-74523b01223c/D155.027%20GA.pdf>, last visited 2022-04-14

⁶⁸<https://noyb.eu/en/edps-sanctions-parliament-over-eu-us-data-transfers-google-and-stripe>, last visited 2022-04-14

⁶⁹<https://noyb.eu/en/update-cnil-decides-eu-us-data-transfer-google-analytics-illegal>, last visited 2022-04-14

⁷⁰<https://help.shopify.com/en/manual/reports-and-analytics/shopify-reports/overview-dashboard>, last visited 2021-12-16

⁷¹<https://woocommerce.com/usage-tracking/>, last visited 2021-12-16

⁷²[Charter of Fundamental Rights of the European Union, 2012], Art. 7

⁷³[Charter of Fundamental Rights of the European Union, 2012], Art. 8

Everyone has the right to respect for his or her private and family life, home and communications.

Legal text 5.1: Article 7 of the Charter of Fundamental Rights of the European Union⁷²

1. Everyone has the right to the protection of personal data concerning him or her.
2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.
3. Compliance with these rules shall be subject to control by an independent authority.

Legal text 5.2: Article 8 of the Charter of Fundamental Rights of the European Union⁷³

5.3.1 Cookies

Based on the findings of sections 4.3.1 and 4.3.3 as well as section 3.3.1, it is given that nearly all of Austria's Top 500 enterprises use cookies. Web scraping results and survey responses have shown that first- as well as third-party cookies are utilised.

Use of cookies

Generally speaking, there is no law prohibiting the use of cookies. However, there are several legal texts regulating such use. Aforementioned Article 5(3) of the Directive on privacy and electronic communications, amended in the E-privacy Directive, and its national transposition in Austria within the TKG21 must be considered and if the data involved is considered to be personal data, the GDPR and DSG are relevant as well.

Applicability of Article 5(3) of the Directive on privacy and electronic communications:

Member States shall ensure that the storing of information, or the gaining of access to information already stored, in the terminal equipment of a subscriber or user is only allowed on condition that the subscriber or user concerned has given his or her consent, having been provided with clear and comprehensive information, in accordance with Directive 95/46/EC, inter alia, about the purposes of the processing. This shall not prevent any technical storage or access for the sole purpose of carrying out the transmission of a communication over an electronic communications network, or as strictly necessary in order for the provider of an information society service explicitly requested by the subscriber or user to provide the service.

Legal text 5.3: Article 5(3) of Directive on privacy and electronic communications⁷⁴

As mentioned in section 2.3.1, cookies are placed in a website visitors browser storage. Browsers are (normally) installed on terminal equipment, making the storing or accessing of already placed cookies "*storing of information, or the gaining of access to information*

⁷⁴[Directive on privacy and electronic communications, 2002], Art. 5(3)

already stored, in the terminal equipment of a subscriber or user". This assumption is supported by Recital 25 of the Directive on privacy and electronic communications. In conclusion, Article 5(3) of the Directive on privacy and electronic communications is applicable to cookies. However, as part of a directive it is not directly applicable, it has to be transferred into national law first.

Applicability of Section 165 Paragraph 3 of the Telekommunikationsgesetz 2021:

The Austrian implementation of Article 5(3) of the Directive on privacy and electronic communications impacts its applicability to cookies. Based on the first part of the first sentence, the paragraph (see legal text 5.4 for the German original text) is only applicable to providers of public communications services or providers of an information society service as defined in the ECG, Section 3 Subparagraph 1 ("*Betreiber öffentlicher Kommunikationsdienste und Anbieter eines Dienstes der Informationsgesellschaft im Sinne des § 3 Z 1 E-Commerce-Gesetz, BGBl. I Nr. 152/2001*"). The definition of the former can be combined from a number of subparagraphs in Section 4 of the TKG21, but is not relevant to this thesis as all of the investigated enterprises fall into the latter category. For the full English text defining the latter category see legal text 5.5.

Betreiber öffentlicher Kommunikationsdienste und Anbieter eines Dienstes der Informationsgesellschaft im Sinne des § 3 Z 1 E-Commerce-Gesetz, BGBl. I Nr. 152/2001, sind verpflichtet, den Nutzer oder Benutzer darüber zu informieren, welche personenbezogenen Daten er verarbeiten wird, auf welcher Rechtsgrundlage und für welche Zwecke dies erfolgt und für wie lange die Daten gespeichert werden. Eine Ermittlung dieser Daten ist nur zulässig, wenn der Nutzer oder Benutzer seine Einwilligung dazu aktiv und auf Grundlage von klaren und umfassenden Informationen erteilt hat. Dies steht einer technischen Speicherung oder dem Zugang nicht entgegen, wenn der alleinige Zweck die Durchführung der Übertragung einer Nachricht über ein Kommunikationsnetz ist oder, wenn dies unbedingt erforderlich ist, damit der Anbieter eines Dienstes der Informationsgesellschaft, der vom Nutzer oder Benutzer ausdrücklich gewünscht wurde, diesen Dienst zur Verfügung stellen kann. Der Nutzer ist auch über die Nutzungsmöglichkeiten auf Grund der in elektronischen Fassungen der Verzeichnisse eingebetteten Suchfunktionen zu informieren. Diese Information hat in geeigneter Form, insbesondere im Rahmen Allgemeiner Geschäftsbedingungen und spätestens bei Beginn der Rechtsbeziehungen zu erfolgen.

Legal text 5.4: Section 165 Paragraph 3 of the Telekommunikationsgesetz 2021 (German original)⁷⁵

The second part of the first sentence together with the second sentence of Section 165 Paragraph 3 of the TKG21 is limiting the applicability to the processing, storage and collection of personal data. It could be argued that the storage of cookies within a user's browser does not fall under those terms as long as the cookie itself does not contain personal data or is being considered personal data (find more detailed information on

⁷⁵[Telekommunikationsgesetz 2021, BGBl. I Nr. 190/2021, 2021], Sect. 165 Para. 3

this distinction further below). The accessing, especially the accessing by third parties, of cookies containing personal data or being considered personal data can definitely be subsumed under those terms as it provides additional insight into the users behaviour, e.g. browsing patterns or reoccurring website visits. This would also allow the storage of not strictly necessary cookies which contain neither personal data nor are considered to be personal data without consent as long as they would not be processed until a user has given their consent. However, such behaviour would make it harder for a user to determine if their data is being processed or not.

The second half of the first sentence of Section 165 Paragraph 3 of the TKG21 contains an obligation for the provider to make certain information about cookies containing or being considered personal data known to users and subscribers. Details on said information will be further discussed, together with the content of the second sentence, in the context of informed consent further below.

"information society service" shall mean a service normally provided in return for consideration electronically by distance selling at the individual retrieval of the recipient (§ 1 para 1 sub-para 2 of the Notification Act of 1999), particularly the online marketing of goods and services, online information offers, online advertising, electronic search engines and data enquiry options as well as services which transmit information via an electronic network and provide access to such a network or store the information of a user;

Legal text 5.5: Section 3 Subparagraph 1 of E-Commerce Gesetz in its official English version⁷⁶

There is a distinction made in German when stating who has to be informed by the provider ("Nutzer oder Benutzer"), which can not be transferred to English directly. Both terms are translated by the same English word: "user". The definition of "Nutzer"⁷⁷ follows logically the definition of "subscriber" given in Recital 12 and 13 of the Directive on privacy and electronic communications. While the definition of "Benutzer"⁷⁸ is nearly a word by word translation of the definition of "user" of Article 2 Subparagraph a of the Directive on privacy and electronic communications. It contains one little difference: The TKG21 defintion misses the word "natural" in front of person, and therefore leaves room for interpretation whether a user could also be a legal person.

Applicability of Article 5(1) Paragraph e of the General Data Protection Regulation:
As Section 165 Paragraph 3 of the TKG21 limits its applicability to personal data, Article 5(1) Paragraph e of the GDPR is applicable to all data collected under the aforementioned paragraph. Based on the wording of Article 5(1) Paragraph e of the GDPR (see legal text 5.6), it is possible to store personal data longer than necessary for the intended

⁷⁶https://www.ris.bka.gv.at/Dokument.wxe?Abfrage=Erv&Dokumentnummer=ERV_2001_1_152, last visited 2021-12-01

⁷⁷[Telekommunikationsgesetz 2021, BGBl. I Nr. 190/2021, 2021], Sect. 4 Subpara. 13

⁷⁸[Telekommunikationsgesetz 2021, BGBl. I Nr. 190/2021, 2021], Sect. 160 Para. 3 Subpara. 2

Diese Webseite verwendet Cookies von Google um Zugriffe zu analysieren. Durch die Nutzung dieser Webseite erklären Sie sich damit einverstanden. [Accept](#)

Figure 5.1: Example: Consent notice with insufficient information and no proper consent option⁸⁰

purposes, as long as it anonymised as soon as it has fulfilled its purpose. It might not be the simplest task to determine when a cookies has reached that point.

1. Personal data shall be:

[...]

(e) kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject ('storage limitation');

Legal text 5.6: Article 5(1) Paragraph e of the General Data Protection Regulation⁷⁹

The paragraph provides exemptions allowing longer storage only for "*archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1)*".

Consent to cookies

Although Article 5(3) of the Directive on privacy and electronic communications makes no such distinction, Section 165 Paragraph 3 of the TKG21 requires active and informed consent by the subscriber or user only if cookies contain personal data. This section should clarify why consent notices, like the one in fig. 5.1, do not fulfill the requirement for active and informed consent.

Personal data:

"Personal data" ("personenbezogene Daten") is neither explicitly defined in the TKG21 nor in Directive on privacy and electronic communications. As processing of personal data is regulated by the GDPR, the definition in legal text 5.7 applies. Recital 30 of the GDPR explicitly mentions cookie identifiers and the possibility of them being used to identify a natural person. In conclusion, legal text 5.4 applies to cookies identifying a natural person and all other cookies containing personal data as defined under the GDPR.

Active consent:

In its second sentence Section 165 paragraph 3 of the TKG21 allows the collection of

⁷⁹[General Data Protection Regulation, 2016], Art. 5(1) Para. e

⁸⁰<https://schmidholding.com/>, last visited 2021-12-13

⁸¹[General Data Protection Regulation, 2016], Art. 4 Subpara. 1

'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

Legal text 5.7: Article 4 Subparagraph 1 of the General Data Protection Regulation⁸¹

personal data only when the subscriber or user has consented actively and based on clear and comprehensive information. The obligation to obtain consent of the person concerned is in line with Article 8 Paragraph 2 of the CFR.

'consent' of the data subject means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her;

Legal text 5.8: Article 4 Subparagraph 11 of the General Data Protection Regulation⁸²

Active consent in case of cookies has been clarified drastically by the preliminary ruling of the ECJ in the case of "Planet 49".⁸³ The Directive on privacy and electronic communications defines consent by referring to Directive 95/46/EC, which has since been repealed by the GDPR. The GDPR's full definition of consent can be found in legal text 5.8. Additional information is provided in Recital 32 of the GDPR. According to Recital 17, "*consent may be given by any appropriate method enabling a freely given specific and informed indication of the user's wishes, including by ticking a box when visiting an Internet website.*". In its preliminary ruling of 1 October 2021 the ECJ clarified that this consent has to be given actively⁸⁴ and states that consent obtained by pre-ticked boxes does not constitute "*freely given, specific, informed and unambiguous' indication of the data subject's wishes in the form of a statement or of 'clear affirmative action*".⁸⁵ Recital 32 of the GDPR also supports this statement.⁸⁶ Figure 5.2 shows a consent notice with an active toggle for statistics cookies, the toggle was already activated when the consent notice appeared. If a user would click the "Accept"-button ("Zustimmen") of that specific consent notice, their consent for statistics cookies would not be considered "*freely given, specific, informed and unambiguous*".

Informed consent:

Sentence 2 of Section 165 Paragraph 3 of the TKG21 states consent has to be given based on clear and comprehensive information ("*auf Grundlage von klaren und umfassenden*

⁸²[General Data Protection Regulation, 2016], Art. 4 Subpara. 11

⁸³[Planet49 (C-673/17), 2019]

⁸⁴[Planet49 (C-673/17), 2019], Para. 52

⁸⁵[Planet49 (C-673/17), 2019], Para. 61

⁸⁶[Planet49 (C-673/17), 2019], Para. 62

⁸⁷<https://www.hogast.at/>, last visited 2021-12-13

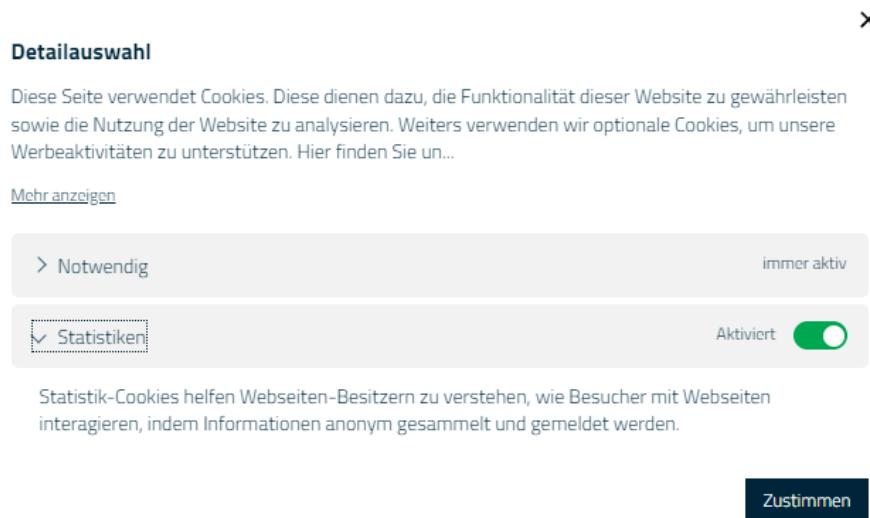


Figure 5.2: Example: Consent notice with preticked box⁸⁷

Informationen"). This relates to the similarly worded passage in Article 5(3) of Directive on privacy and electronic communications ("*having been provided with clear and comprehensive information*"). The Article also states this information must be given "*in accordance to Directive 95/46/EC, inter alia*" ("*inter alia*" meaning "*among other things*"), Section 165 Paragraph 3 of the TKG21 does not include any obligations on this matter.

As mentioned before, the second half of the first sentence of Section 165 Paragraph 3 of the TKG21 lists information, which has to be disclosed to the subscriber or user by providers of public communications services or providers of an information society service. Information provided must include what kind of personal data is processed, the legal basis as well as the purpose for which it is processed and how long it is stored. This list is closely related to the information required by Article 13 of the GDPR. As Section 165 Paragraph 3 of the TKG21 is limited to the collection of personal data, the aforementioned article is applicable anyway.

Figure 5.3 shows the information provided by a company which does not even use a consent notice to obtain consent for their analytics cookies. The information provided is missing details and, overall, can be considered insufficient. In contrast, fig. 5.4 provides all required information in a clear and well-formed manner.

There is another obligation for disclosure mentioned in the last two sentences of Section 165 Paragraph 3 of the TKG21. The first of those two sentences enforces Article 12(1) of the Directive on privacy and electronic communications, in particular the necessity of informing a subscriber about the "*search functions embedded in electronic versions of*"

⁸⁸<https://www.oebb.at/en/rechtliches/nutzungsbedingungen>, last visited 2021-12-13

⁸⁹<https://www.transped.at/eucookie/eucookie>, last visited 2021-12-13

6. Web analysis service

- Our websites as well as our digital communication with our customers (e.g. our Newsletter) use Matomo. This is a so-called web analysis service. Matomous so-called "cookies". These are text files which are stored on your computer and allow us to analyse your usage of the web pages. For this purpose, the usage information generated by the cookie (including your truncated IP address) is transferred to our server and stored for usage analysis purposes. This helps us in optimising our web pages. During this procedure, your IP address is immediately anonymised, so that you remain anonymous to us.
 - The information generated by cookies on the usage of our websites is not passed on to third parties.
 - You can prevent the use of cookies through according settings in your browser software. This may, however, result in you not being able to fully use all functions provided by our websites.
 - If you do not agree to the storage and analysis of data related to your visit and usage of our websites, you can object to such storage and usage by clicking below. In this case, a so-called opt-out cookie will be stored in your browser. As a result, Matomo will collect no session data. We expressly draw your attention to the fact that if you delete your cookies, this will also result in your opt-out cookie being deleted. It would therefore then have to be re-activated by you.
- Here, you can decide whether or not a unique web analysis code is allowed to be stored in your browser to allow us the collection and analysis of different statistical data. If you do not agree, please click on the following link in order to set the Piwik deactivation cookie in your browser. Your visit to this web page is currently being recorded by the Piwik web analysis. Click here for your visit to no longer be recorded.

You are not being tracked since your browser is reporting that you do not want to. This is a setting of your browser so you won't be able to opt-in until you disable the 'Do Not Track' feature.

Figure 5.3: Example: Insufficient cookie information⁸⁸

Übersicht verwendeter Cookies

GRUPPE: ESSENTIELL

Essentielle Cookies werden für grundlegende Funktionen der Website benötigt. Dadurch wird sichergestellt, dass die Website ordnungsgemäß funktioniert.
Name: dzEuCookieConsent
Beschreibung: Wird verwendet um die vom Benutzer ausgewählte Cookie-Einstellungen zu speichern.
Speicherdauer: 365 Tage
Provider: Gilbert Zimmermann, DataZimmermann
Name: dzLanguageSelection
Beschreibung: Wird verwendet um die vom Benutzer ausgewählte Sprache zu speichern.
Speicherdauer: 365 Tage
Provider: Gilbert Zimmermann, DataZimmermann

GRUPPE: MARKETING

Marketingcookies umfassen Tracking- und Statistikcookies.
Name: _ga, _gat, _gid
Beschreibung: Diese Cookies werden von Google Analytics verwendet, um verschiedene Arten von Nutzungsinformationen zu sammeln, einschließlich persönlicher und nicht-personenbezogener Informationen. Weitere Informationen finden Sie in den Datenschutzbestimmungen von Google Analytics unter policies.google.com/privacy. Gesammelte nicht personenbezogene Daten werden verwendet, um Berichte über die Nutzung der Website zu erstellen, die uns helfen, unsere Websites / Apps zu verbessern.
Speicherdauer: Unterschiedlich: 2 Jahre, 6 Monate oder kürzer.
Provider: Google Analytics
Name: _fbp
Beschreibung: Auf diesen Seiten werden Social media Plug-ins eingebunden, um Artikel in Sozialen Netzwerken, beispielsweise Facebook, Twitter, empfehlen und teilen zu können. Hierfür setzen wir ein zweistufiges Verfahren ein. Daten an Dritte werden erst überfragen, wenn Nutzer auf eines der in der Social Media Leiste angezeigten Icons klicken. Wir haben dabei weder Einfluss auf Cookies, die von Facebook, Twitter etc. gesetzt werden, noch hat wir Zugriff darauf.
Speicherdauer: 3 Monate
Provider: Facebook

Figure 5.4: Example: Well-defined cookie information⁸⁹

directories. A telephone book would be such a directory. The placement of this obligation reduces readability - I had to refer to the German TTDSG to find a similar worded passage⁹⁰ to understand the phrase: "*in elektronischen Fassungen der Verzeichnisse eingebetteten Suchfunktionen*". Especially, because the word directory ("Verzeichnis") is used in more than one context within the TKG21. However, there is no use case for this obligation in regard to cookies and therefore, it will not be discussed any further.

Conditions for consent:

Providers are required to document a data subject's consent under Article 7(1) of the GDPR. The article states that providers "*shall be able to demonstrate that the data subject has consented to processing*". It does, however, not define a specific way, in which such consent must be stored. In case of cookies, most CMPs store consent itself within a cookie. Widespread examples are, the "cookielaawinfo-checkbox-necessary" cookie by the GDPR Cookie Consent plugin or "CookieConsent" by Cookiebot, which were both found in the Top 25 cookies in section 4.3.1. While the first saves consent separately for every category of cookies (the consent for "necessary" is mostly pre-ticked and therefore already present on page load), the latter saves consent in general.

Under Article 7(3) of the GDPR, it "*shall be as easy to withdraw as to give consent*.". A lot of companies provide a single button to accept all cookie, but do not offer the same option when it comes to rejecting all cookies. The European non-profit organisation "noyb" found that 81% of all websites in their data set did not allow cookie rejection via the first layer of the cookie consent banner or pop-up, instead these options are hidden within sub-menus.⁹¹ This is a so-called "dark pattern", used to nudge users into doing something they would likely prefer not doing by making alternative options less convenient. In 2018, the Norwegian Consumer Council published a report on dark patterns, stating: "*If the aim is to lead users in a certain direction, making the process toward the alternatives a long and arduous process can be an effective dark pattern. This relates to the issue of defaults, since the default setting clearly is the easiest option for the user. It is however also easier to see and act on some designs or colours than others, and making some buttons or options more salient may also affect our choices*"⁹² All those practices are clearly in violation of Article 7(3) of the GDPR.

The example in fig. 5.5 shows a rejection option on the first layer of a consent notice without the utilisation of any dark patterns. In fig. 5.6 the "Accept"-button ("Speichern") is clearly more present than the other options, including the rejection option, due to its colour. This consent notice uses one of the aforementioned dark patterns.

⁹⁰[Telekommunikation-Telemedien-Datenschutz-Gesetz, BGBl. I S. 1982 idF I S. 3544, 2021], Sect. 17 Para. 1

⁹¹<https://noyb.eu/en/noyb-aims-end-cookie-banner-terror-and-issues-more-500-gdpr-complaints>, last visited 2021-12-07

⁹²[Norwegian Consumer Council, 2018], p. 19

⁹³<https://www.linde-gas.at/de/index.html>, last visited 2021-12-13

⁹⁴<https://www.bundesforste.at/>, last visited 2021-12-13

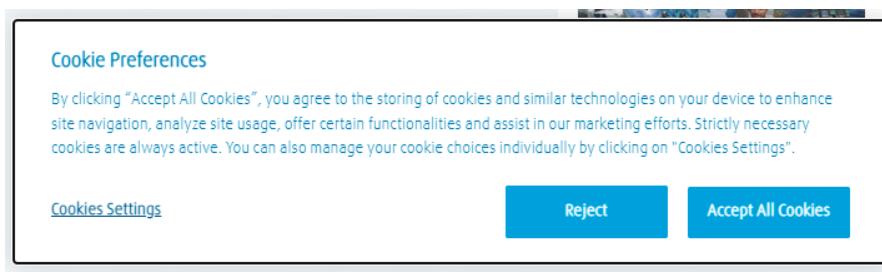


Figure 5.5: Example: Rejection option on the first layer of a cookie consent pop-up⁹³



Figure 5.6: Example: Utilisation of a dark pattern⁹⁴

Strictly necessary cookies

The last sentence of Article 5(3) of the Directive on privacy and electronic communications states two exemptions. The first exemption allows technical storage of or access to cookies needed for *carrying out the transmission of a communication over an electronic communications network*. The second also allows storage of or access to cookies, but only for cookies which are strictly necessary to provide explicitly requested services to a website visitor. The assessment of what is "strictly necessary" can be difficult - if there is any doubt, it is recommended to ask for the user's consent. An example for obviously false categorisation of cookies can be seen in fig. 5.7.

Additionally, strictly necessary cookies can only be used for the purpose for which they are strictly necessary; a user would still have to consent to any other use of the same cookie. It is also recommended to still provide information about strictly necessary cookies to the user.⁹⁶ If a CMP is used to provide this information, the provider should make sure the CMP knows all utilised cookies to avoid situations like the description ("Beschreibung") in fig. 5.8.

The respective Austrian legislation, Section 165 Paragraph 3 of the TKG21, contains

⁹⁵<https://www.marienhuette.at/>, last visited 2021-12-13

⁹⁶<https://gdpr.eu/cookies/>, last visited 2021-12-10

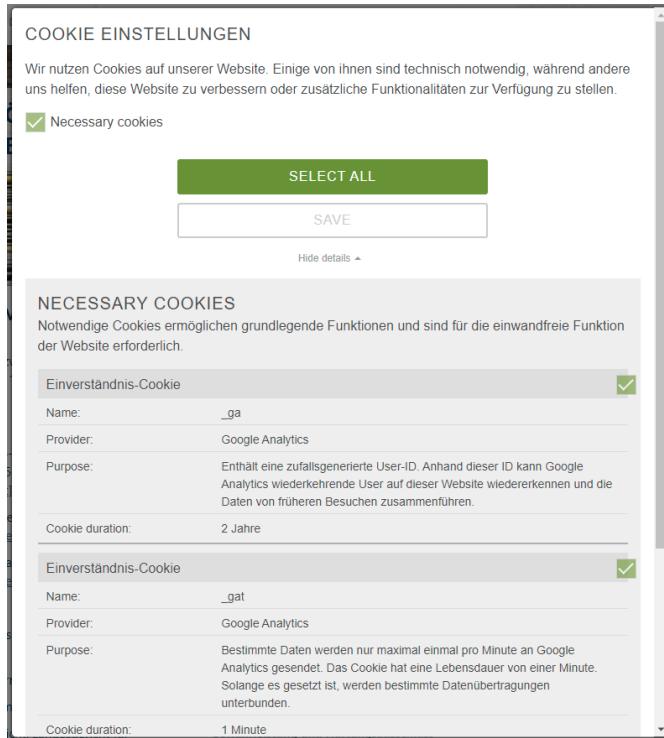


Figure 5.7: Example: False cookie categorisation⁹⁵

a literal translation of the last sentence of Article 5(3) of the Directive on privacy and electronic communications.

Summary of legal requirements

Based on the preceding sections, it can be concluded that:

- Strictly necessary cookies can be stored and accessed under Austrian and EU law without consent by the user
- Cookies containing or being considered personal data are subject to the following restrictions:
 - the provider (as defined in Section 165 Paragraph 3) must disclose information on what kind of personal data is processed, the legal basis as well as the purpose for processing it and how long it is stored
 - the user or subscriber has given active consent to the collection of their personal data on the basis of aforementioned information

⁹⁷<https://www.dhl.com/at-de/home.html>, last visited 2021-12-13

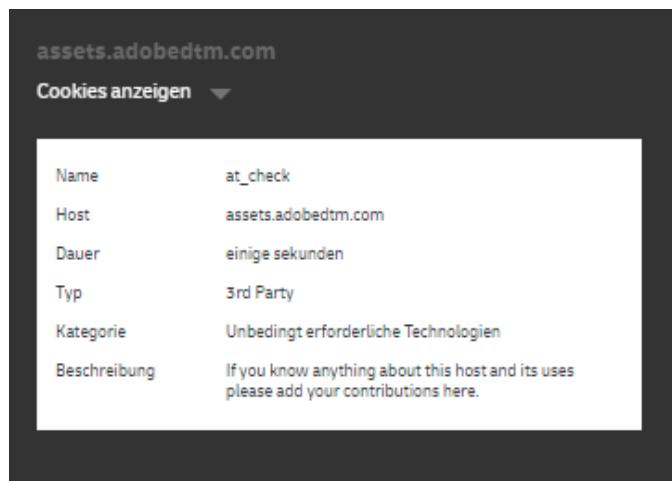


Figure 5.8: Example: Missing information on strictly necessary cookie⁹⁷

- consent can be rejected or withdrawn as easily as it can be given by the user or subscriber
- collected data is stored not longer than necessary for the purposes

Enterprises with more than one place of jurisdiction within the EU should consider that the limited obligation to acquire consent only for cookies containing or being considered personal data, is solely a result of Section 165 Paragraph 3 of the TKG21. Such a limitation is not supported by EU law. In contrary, the ECJ explicitly stated that *"Article 2(f) and Article 5(3) of Directive 2002/58, read in conjunction with Article 2(h) of Directive 95/46 and Article 4(11) and Article 6(1)(a) of Regulation 2016/679, are not to be interpreted differently according to whether or not the information stored or accessed on a website user's terminal equipment is personal data within the meaning of Directive 95/46 and Regulation 2016/679."*⁹⁸ Therefore, those enterprises should look at the national transposition of Article 5(3) of the Directive on privacy and electronic communications in the respective Member States (for example Section 25 of the TTDSG for Germany).

Special cases

Session cookies:

The results from section 4.3.1 show that a relatively high number among the most frequently found cookies on websites provided by Austria's Top 500 enterprises are session cookies. In most cases session cookies are part of the category of strictly necessary cookies. They can, for example, be used to keep users logged in across different pages of a website,

⁹⁸[Planet49 (C-673/17), 2019], Para. 71

or protect a webserver from illegal activities, like DoS attacks.⁹⁹ In this case, they fall under the consent exemption, which has already been discussed in section 5.3.1.

In some cases session cookies combined with other information (like browsed pages within a website, which are only accessible to a specific user) or linked with other data provided by the user, could be "*used to create profiles of [...] natural persons and identify them.*"¹⁰⁰ In these cases, the rules for not strictly necessary cookies, see section 5.3.1, apply.

Cookie-based third-party tracking:

In addition to the rules applicable to genuine first-party cookies established in section 5.3.1, providers have to disclose "*whether or not third parties may have access to those cookies.*"¹⁰¹

The results from section 4.3 show that (Referred) Analytics tracking, Vanilla tracking as well as Personal tracking as defined by Lerner et al.¹⁰² is present on websites provided by Austria's Top 500 enterprises.

Cookies set in first-party context which are later leaked to third parties are considered Analytics tracking by Lerner et al.¹⁰² For example, in section 4.3.1 a number of cookies set by Google Analytics in the first-party context of other domains were found. Google itself provides a page informing about possible data sharing settings of its Analytics service. Google states: "Regardless of your data sharing settings, your Analytics data may also be used only insofar as necessary to maintain and protect the Analytics service."¹⁰³ The settings provided go so far as to allow Google full access to the collected data. In this case, "*Google is, for GDPR purposes, an independent controller of such data.*"¹⁰³ However, Google tries to transfer the obligation to request consent by the user entirely to its customer which is using the Analytics service.¹⁰⁴ According to the ECJs preliminary ruling on "Fashion ID", the first party is only required to inform the user and acquire consent by the user for any "*processing of personal data in respect of which that operator [Annot.: first party] determines the purposes and means.*"¹⁰⁵ If any of the other data sharing settings are chosen, Google remains a data processor under the GDPR¹⁰⁶ and the first party must enter into a data processing agreement with Google in accordance with Article 28(3) of the GDPR.

The regulation of Vanilla Tracking is illustrated by the following example: Third-party content embedded as an iframe might be able to set third-party cookies. The required

⁹⁹https://cheatsheetseries.owasp.org/cheatsheets/Session_Management_Cheat_Sheet.html, last visited 2021-12-04

¹⁰⁰[General Data Protection Regulation, 2016], Rec. 30

¹⁰¹[Planet49 (C-673/17), 2019], Para. 81

¹⁰²[Lerner et al., 2016], p. 1001

¹⁰³<https://support.google.com/analytics/answer/1011397>, last visited 2021-12-11

¹⁰⁴<https://support.google.com/analytics/answer/9012600>, section 4.3, last visited 2021-12-11

¹⁰⁵[Fashion ID (C-40/17), 2019], Para. 106

¹⁰⁶<https://support.google.com/analytics/answer/6004245#zippy=%2Cgoogle-analytics-under-the-general-data-protection-regulation-gdpr>, last visited 2021-12-11

information (purpose, storage duration, etc.) about those cookies has to be disclosed to the user and they must give their consent before any third-party cookies can be set.¹⁰⁷ Whenever personal data is transmitted to the third party in such a setup, both parties are considered to be "controllers" in accordance with the GDPR and jointly responsible for the processing of user's personal data.¹⁰⁸ As mentioned above, the first party has limited obligations regarding information and consent by the user in this case. Meaning, for all processing purposes and means not determined by the first party, the obligation to inform and acquire consent lies with the third-party involved. Such a constellation is regulated by Article 26 of the GDPR and requires a joint controller agreement by both parties.

Personal tracking, e.g. via a social plugin like the Facebook "Like"-Button, behaves similarly to Vanilla tracking. The same set of rules applies.

Possible fines

Penalties, especially regarding violations of the GDPR, are dependent on the circumstances of an individual case and complex in nature. Therefore, the following list should be considered exemplary.

Violation of Section 165 Paragraph 3 of the TKG21:

The TKG21 established a fine of 50,000€, or in case of uncollectability 6 weeks of imprisonment, for not providing users or subscribers with the information defined in Section 165 Paragraph 3.¹⁰⁹

Due to the fact that the applicability of Section 165 Paragraph 3 is limited to the collection of personal data, penalties for processing such data without consent are not explicitly provided by the TKG21. Processing personal data without consent (or other legal basis) is already punishable in accordance with the GDPR.

Violation of the GDPR:

Penalties for violating the GDPR are established in its Article 83. Depending on the individual case, fines can be issued instead of or in addition to measures in accordance with Article 58 of the GDPR. E.g. when personal data is processed without consent, where no other legal basis for processing is given, fines can be "*up to 20 000 000 EUR, or in the case of an undertaking, up to 4% of the total worldwide annual turnover of the preceding financial year, whichever is higher*".¹¹⁰ The same applies when the withdrawal of consent is not as easy as the act of consenting or whenever the data subject's rights (Articles 12 to 22 of the GDPR) are violated.

Infringements of the "*the obligations of the controller and the processor pursuant to Articles 8, 11, 25 to 39 and 42 and 43 [Annot.: of the GDPR]*", for example not keeping records of processing activities, can be fined with "*up to 10 000 000 EUR, or in the*

¹⁰⁷ [Directive on privacy and electronic communications, 2002], Art. 5(3)

¹⁰⁸ [Fashion ID (C-40/17), 2019], Para. 85

¹⁰⁹ [Telekommunikationsgesetz 2021, BGBl. I Nr. 190/2021, 2021] Sect. 188 Para. 4 Subpara. 24

¹¹⁰ [General Data Protection Regulation, 2016], Art. 83(5) Subpara. a

case of an undertaking, up to 2% of the total worldwide annual turnover of the preceding financial year, whichever is higher".¹¹¹

5.3.2 HMTL5 localStorage

Section 4.3.3 showed that nearly 60% of all websites provided by Austria's Top 500 enterprises use the HTML5 localStorage. The legal implications of tracking-related utilisation of HTML5 localStorage are discussed below.

Use of HTML5 localStorage

As with cookies, there is no law generally prohibiting the use of HTML5 localStorage. Given the technical similarity and near identical usage of cookies and information stored in HTML5 localStorage, I would argue that the same legislation applies.

Applicability of Article 5(3) of the Directive on privacy and electronic communications:
Based on the HTML5 localStorage being embedded in a user's browser (for more information see section 2.3.1), storing and accessing information there can be considered "*storing of information, or the gaining of access to information already stored, in the terminal equipment of a subscriber or user*",¹¹² thus making Article 5(3) of the Directive on privacy and electronic communications applicable.

Applicability of Telekommunikationsgesetz 2021, Section 165 Paragraph 3:

Based on the applicability of Article 5(3) of the Directive on privacy and electronic communications, Section 165 Paragraph 3 of the TKG21 must be applicable as well. The same limitations already mentioned in section 5.3.1 have to be considered.

Applicability of Article 5(1) Paragraph e of the General Data Protection Regulation:
Due to the technical nature of HTML5 localStorage, all contained data is persistent until a browser's localStorage has been cleared by the user or by the website itself. As Article 5(1) Paragraph e of the General Data Protection Regulation requests that data should only be stored for as long as it is needed for the purpose it was collected for, all data stored in HTML5 localStorage serve a specific purpose warranting persistent storage to comply with said article.

Applicability of Article 15 to 22 of the General Data Protection Regulation:

As Section 165 Paragraph 3 of the TKG21 is limited to personal data, Article 15 to 22 of the GDPR is applicable.

The limitations on data manipulation mentioned above are making it close to impossible for a provider to delete or alter stored data objects in HMTL5 localStorage upon request by the user. Therefore, it is unclear if a provider would be able to uphold all rights of a data subject under Article 15 to 22, but especially Article 17 (1) Subparagraph b, of the GDPR when personal data is stored in the HTML5 localStorage.

¹¹¹[General Data Protection Regulation, 2016], Art. 83(4) Subpara. a

¹¹²[Directive on privacy and electronic communications, 2002], Art. 5(3)

atlassian-analytics.confluence.lock	Extranet Wikis Public Wikis	Used for statistics only, in order to store statistics about Confluence usage	localStorage, persistent data
atst.healthcheck.sensors.page-protocols	Extranet Wikis Public Wikis	Keeps Confluence analytics disabled	localStorage, persistent data
confluence.sidebar.width	Extranet Wikis Public Wikis	Remembers the width of the sidebar	localStorage, persistent data
dashboard.route.last	Extranet Wikis Public Wikis	Remembers preference for information displayed on the user's dashboard	localStorage, persistent data

Figure 5.9: Example: Information on HTML5 localStorage objects¹¹⁴

Consent to data being stored in HTML5 localStorage

The same set of rules that apply to cookies also applies to data being stored in HTML5 localStorage. However, very few companies provide any information on data being placed in HTML localStorage. A likely cause is that consent notices created by common CMPs often lack the capability to easily include HTML5 localStorage.¹¹³ Figure 5.9 shows the information provided on HTML5 localStorage objects provided on the website of the European Commission.

Another issue of utilising HTML5 localStorage might arise from the method used by many CMPs to store a user's consent. As mentioned in section 5.3.1, consent is often stored by means of cookies. This consent must be demonstrated by the provider if necessary. However, cookies are normally set to expire. In case the cookie, in which consent was stored, expires before HTML5 localStorage is wiped, the provider loses the ability to demonstrate that consent was given.

Strictly necessary data storage in HTML5 localStorage

Exemptions discussed in section 5.3.1 with respect to the above mentioned legislation apply to HTML5 localStorage as well. However, there is little to no information that actually needs to be stored for longer than one browser session (storage duration limitation of HTML5 sessionStorage) and is bigger than 4 kilobyte (size limitation of cookies). Only

¹¹³<https://www.dataprotectionreport.com/2019/06/nt-analyzer-blog-series-why-so-many-cookie-policies-are-broken-part-i-html5-localstorage/>, last visited 2021-12-13

¹¹⁴https://ec.europa.eu/info/cookies_en, last visited 2021-12-13

if both requirements are met, it can be argued that strictly necessary information has to be stored in HTML localStorage.

Summary of legal requirements

Similarly to application of aforementioned law to cookies, it is concluded that:

- Strictly necessary data can be stored in and accessed from HTML5 localStorage under Austrian and EU law without consent by the user
- Whenever information stored in HTML5 localStorage contains personal data or can be considered as such the following restrictions apply:
 - the provider must disclose information on what kind of personal data is processed, the legal basis as well as the purpose for processing it and how long it is stored
 - the user or subscriber has given active consent to the collection of their personal data on the basis of aforementioned information
 - consent can be rejected or withdrawn as easily as it can be given by the user or subscriber
 - collected data warrants persistent storage (by law it must not be stored longer than necessary for the purposes¹¹⁵ and in case of HTML5 localStorage storage is always persistent)

Again, enterprises with more than one place of jurisdiction within the EU should consider that the limitation of consent to HTML5 localStorage objects containing or being considered personal data is solely a result of Section 165 Paragraph 3 of the TKG21. It is not supported by EU law. Therefore, those enterprises should look at the national transpositions of Article 5(3) of the Directive on privacy and electronic communications in their respective Member State.

Special cases

Third-party data stored in HTML5 localStorage:

A little less than 20% of all objects stored in HTML5 localStorage, which were detected by webXray (see section 4.3.3 for more information), were placed by third parties. Like cookies which are accessible to third parties, information stored in HTML localStorage which can be accessed by third parties must be disclosed to users.

Overall, the similarities between storing and accessing third-party information in the HTML5 localStorage and storing and accessing third-party cookies provide ample basis for the assumption that the same legislation is applicable.

¹¹⁵[General Data Protection Regulation, 2016], Art. 5(1) Subpara. e

Possible fines

As with general applicability of legal texts, also the argumentation for fines closely follows the one for cookies. See section 5.3.1 for more information.

5.3.3 Fingerprinting

There is not a lot of conclusive evidence on fingerprinting methods employed by Austria's Top 500 enterprises. Therefore, this section provides only a very superficial summary on potentially relevant legislation in this matter, followed by two special cases.

Regulation of fingerprinting

No European or Austrian law prohibits any form of fingerprinting. However, in 2014 the Article 29 Working Party (*"the independent European working party that dealt with issues relating to the protection of privacy and personal data until 25 May 2018"*¹¹⁶) provided an opinion stating that Article 5(3) of the Directive on privacy and electronic communications is applicable to fingerprinting when certain criteria are met.¹¹⁷ They also concluded that device fingerprints can constitute personal data.¹¹⁸

Whenever a fingerprint is obtained via storing or accessing already stored information Article 5(3) of the Directive on privacy and electronic communications applies. Additionally, a fingerprint containing personal data or data identifying a natural person (in combination with other data available to the provider) is regulated by the GDPR. Both cases require consent, with the previously discussed exemptions; this and other relevant parts of the mentioned legal texts have already been covered in this thesis. Refer to section 5.3.1 and section 5.3.2 for more information.

Special cases

Browser fingerprinting:

Googles tracking pixel "utm.gif" found in section 4.3.3 uses browser fingerprinting techniques according to Google's own description: *"The HTTP request for any web page contains details about the browser and the computer making the request, such as the hostname, the browser type, referrer, and language. In addition, the DOM of most browsers provides access to more detailed browser and system information, such as Java and Flash support and screen resolution."* There is no precedent which would allow to consider any of those actions accessing already stored information on a user's terminal equipment without doubt. However, as an HTTP request also includes the IP address of the requesting party, the following argumentation about IP address tracking should be of interest.

¹¹⁶https://edpb.europa.eu/about-edpb/more-about-edpb/article-29-working-party_en, last visited 2021-12-14

¹¹⁷[Article 29 Data Protection Working Party, 2014], p. 7 f.

¹¹⁸[Article 29 Data Protection Working Party, 2014], p. 4

IP address tracking:

Two respondents to the survey answered that they use Google Fonts. The CMP provider Usercentrics points out that Google Fonts collects IP addresses from requests sent to them¹¹⁹ and Google itself states that it "*logs records of the CSS and the font file requests*".¹²⁰ Because of the following reasons and the information presented in section 2.3.1 on IP address tracking, it can be assumed that Google can identify a great number of natural persons combining existing information and IP addresses: Google Chromes market share is currently above 60%¹²¹ and Chrome signs users in by default whenever they sign in to a Google service¹²²; Android has a market share of more than 70% on mobile devices¹²³ and a convenient use of Android devices is closely tied to the Google Play Store¹²⁴ as well as the fact that Android prompts every user with a Google login screen on setup. Therefore, the GDPR would be applicable and providers would either have to reason why their legitimate interest¹²⁵ to use Google Fonts outweighs the interests or fundamental rights of their users or ask their users for explicit consent to use it. It is made especially difficult to reason legitimate interest in this case by the fact that the IP address is not only processed by a third party but also sent to the U.S. More details about transferring data to the U.S. and the legal implications of it, can be found in section 5.3.7.

The same argument made above applies to another font library named by survey respondents: "Font Awesome". Font Awesome also states that collected data is stored in the U.S. and in case of their content delivery network even "*worldwide*".¹²⁶

At least in other EU countries with more literal transpositions of Article 5(3) of the Directive on privacy and electronic communications, user consent should always be requested. Fonts are stored by Google Fonts up to a year in a browser's cache¹¹⁹, making its use "*storing of information, or the gaining of access to information already stored*".¹²⁷

5.3.4 Email tracking

This section relies solely on survey responses. Three respondents provided information on the newsletter tool their companies are using in order to reach their customers, see section 3.3 for more information.

¹¹⁹<https://usercentrics.com/knowledge-hub/google-fonts-gdpr-compliant/>, last visited 2021-12-05

¹²⁰https://developers.google.com/fonts/faq#what_does_using_the_google_fonts_api_mean_for_the_privacy_of_my_users, last visited 2021-12-09

¹²¹<https://gs.statcounter.com/browser-market-share>, last visited 2021-12-05

¹²²<https://support.google.com/chrome/answer/9159867>, last visited 2021-12-09

¹²³<https://gs.statcounter.com/os-market-share/mobile/worldwide>, last visited 2021-12-05

¹²⁴<https://www.tomsguide.com/news/i-used-android-without-google-here-are-the-pros-and-cons>, last visited 2021-12-09

¹²⁵[General Data Protection Regulation, 2016], Art.6(1) Subpara. f

¹²⁶<https://fontawesome.com/privacy>, last visited 2021-12-09

¹²⁷[Directive on privacy and electronic communications, 2002], Art. 5(3)

Use of newsletters

There is no regulation on newsletters in Austria or the EU as long as they do not constitute commercial communication or are sent for the purpose of direct marketing. However, email addresses are personal data and processing them is governed by the GDPR. Recital 47 of the GDPR states: "*The processing of personal data for direct marketing purposes may be regarded as carried out for a legitimate interest.*" It can therefore be assumed that, under the GDPR, data collected from customers can legally be used for sending out newsletters with the sole purpose of direct marketing, unless otherwise regulated by national law.

Commercial communication:

In Austria, commercial communication is regulated by the ECG, which is the national transposition of the EU's Directive on electronic commerce. The definition of commercial communication in the ECG (legal text 5.10) is a near literal translation of the definition provided by the Directive on electronic commerce (legal text 5.9). Whenever a newsletter can be considered commercial communication, the sender of said newsletter must make sure it is clearly and unambiguously identifiable as commercial communication;¹²⁸ if its unsolicited commercial communication, Section 7 of the ECG even defines a time at which it must be identifiable as such. The sender is also obligated to clearly provide other information in case the newsletter includes promotions or price competitions - all regulated contents of commercial communication are defined in Section 6 Paragraph 1 of the ECG.

"commercial communication": any form of communication designed to promote, directly or indirectly, the goods, services or image of a company, organisation or person pursuing a commercial, industrial or craft activity or exercising a regulated profession. The following do not in themselves constitute commercial communications:

- information allowing direct access to the activity of the company, organisation or person, in particular a domain name or an electronic-mail address,
- communications relating to the goods, services or image of the company, organisation or person compiled in an independent manner, particularly when this is without financial consideration;

Legal text 5.9: Article 2 Subparagraph f of the Directive on electronic commerce¹²⁹

Direct marketing:

The use of newsletters for the purpose of direct marketing is regulated in Austria via the TKG21. The relevant section 174 is the national transposition of Article 13 of the Directive on privacy and electronic communications.

¹²⁸[E-Commerce-Gesetz, BGBl. I 152/2001 idF I 148/2020, 2001], Sect. 6

¹²⁹[Directive on electronic commerce, 2000], Art. 2 Subpara. f

¹³⁰https://www.ris.bka.gv.at/Dokumente/Erv/ERV_2001_1_152/ERV_2001_1_152.html, last visited 2021-12-15

"commercial communication" shall mean advertising and other forms of communication designed to promote, directly or indirectly, the sale of goods and services or the image of a company, except:

- a) information allowing direct access to the activity of the company, e.g. a domain name or an electronic-mail address; as well as
- b) information relating to the goods, services or image of a company, compiled in an independent manner, particularly when this is without financial consideration;

Legal text 5.10: Section 3 Subparagraph 6 of the E-Commerce Gesetz in its official english translation¹³⁰

Section 174 Paragraph 3 of TKG21 prohibits sending electronic mail for the purpose of direct marketing without obtaining explicit consent by the recipient first. Paragraph 5 of this section states, that, in case the sender's identity is masked in any way, the sending of electronic mail is prohibited. Furthermore, the sending of electronic mail is prohibited if Section 6 Paragraph 1 of the ECG is violated, the recipient is asked to visit websites, which violate the aforementioned section or if the sender does not provide an authentic address to which a recipient can send a request to cease such communication. The same set of criteria is mentioned in Article 14(4) of the Directive on privacy and electronic communications.

However, Section 174 Paragraph 4 of the TKG21 provides an exemption for electronic mail meeting all of the following criteria (see legal text 5.11 for the German original):

- the sender has obtained the contact information through a sale or service to their customer
- the electronic mail is sent to promote similar products or services
- the recipient was provided with a clear and unambiguous option to reject the processing of their contact information for such communication and can withdraw consent, free of charge, whenever they receive electronic mail from the sender
- the recipient has not rejected such communications from the start, especially via the opt-out list mentioned in Section 7 Paragraph 2 of the ECG

Use of third-party newsletter tools

Email addresses constitute personal data, the transfer of such into third-party newsletter tools is therefore regulated by the GDPR. The company providing the newsletter tool is considered a "processor" as defined by the GDPR.¹³² The first party transferring personal data must have a contract governing the processing of such with the third party receiving

¹³¹[Telekommunikationsgesetz 2021, BGBl. I Nr. 190/2021, 2021], Sect. 174 Para. 4

¹³²[General Data Protection Regulation, 2016], Art. 4 Subpara. 8

Eine vorherige Einwilligung für die Zusendung elektronischer Post gemäß Abs. 3 ist dann nicht notwendig, wenn

1. der Absender die Kontaktinformation für die Nachricht im Zusammenhang mit dem Verkauf oder einer Dienstleistung an seine Kunden erhalten hat und
2. diese Nachricht zur Direktwerbung für eigene ähnliche Produkte oder Dienstleistungen erfolgt und
3. der Empfänger klar und deutlich die Möglichkeit erhalten hat, eine solche Nutzung der elektronischen Kontaktinformation bei deren Erhebung und zusätzlich bei jeder Übertragung kostenfrei und problemlos abzulehnen und
4. der Empfänger die Zusendung nicht von vornherein, insbesondere nicht durch Eintragung in die in § 7 Abs. 2 E-Commerce-Gesetz genannte Liste, abgelehnt hat.

Legal text 5.11: Section 174 Paragraph 4 of the Telekommunikationsgesetz 2021¹³¹

the data.¹³³ The data subject must have been informed that the first party will disclose their personal data to the third party at the time the personal data was collected.¹³⁴

Data collected by newsletter tools:

It is assumed, that all three newsletter tools mentioned by respondents of the survey collect additional data on newsletter recipients. The assumption regarding the type of collected data is based on information provided in the respective privacy policies. It is possible, that customers can further configure the types of data collected by the newsletter tool and that Austria's Top 500 enterprises collect less or additional data via such tools.

The data mentioned by the company artegic, which is assumed to use their in-house newsletter tool ELAINE, include among other things "*data of the mobile device used*", "*location data based on the IP address*" and "*recommendation via social networks such as Facebook or Twitter*"¹³⁵.

Emarsys states: "*For analysis purposes we link your personal data and the web beacons to your email address and an individual ID. Links received in the newsletter also include this ID. We use these data to create a user profile in order to tailor the newsletter to your personal interests. We track when you read our newsletters and which links you click in them, and we infer your personal interests from this information. We link this information to actions you take on our website. This processing is based on our legitimate interests (Art. 6 (1) lit. f GDPR) in order to provide you with a better usage experience.*"¹³⁶ It is arguable that the fundamental rights and freedoms of the data subject override the "legitimate interest" by the newsletter company in this case, but to the best of my knowledge no legal precedent supporting this assumption exists to date.

¹³³[General Data Protection Regulation, 2016], Art. 28(3)

¹³⁴[General Data Protection Regulation, 2016], Art. 13(1) Subpara. e

¹³⁵<https://www.artegic.com/privacy/>, last visited 2021-12-15

¹³⁶<https://emarsys.com/privacy-policy/>, last visited 2021-12-15

Eworx declare their use of their in-house solution mailworx and state that they track time of email delivery, time when the email was opened, for how long the email was open, IP address of the recipient at the time the email was opened, details on the recipient's email program, links clicked within the email and the time of each click.¹³⁷

All data gathered by the newsletter company can be considered personal data whenever it can be directly linked to an email address of a natural person. If any of that data is used to automatically target users with specific marketing campaigns, it constitutes profiling as defined by the GDPR (see legal text 5.12). Profiling must be disclosed to the data subject at the time their personal data is collected.¹³⁸

'profiling' means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements;

Legal text 5.12: Article 4 Subparagraph 4 of the General Data Protection Regulation¹³⁹

Summary of legal requirements

Based on the preceding sections there are three sets of requirements, depending on the type and distribution model of the respective newsletter. Those sets of requirements are not exclusive, meaning, depending on the newsletter, two or even all three of them could apply at once. As personal data is processed in all of those cases, the processing must be lawful under Article 6 of the GDPR and any profiling must be disclosed to the data subject.

Additionally, the following must be considered:

- If the newsletter constitutes commercial communication under the ECG:
 - all information defined in Section 6 Paragraph 1 of the ECG must be provided in a clear and unambiguous manner
 - in case the communication is unsolicited
 - * the newsletter must be clearly and unambiguously identifiable as unsolicited commercial communication when it is received by the addressee
- If the newsletter is sent for the purpose of direct marketing:
 - the recipient must have given explicit consent except when the newsletter is covered by the exemption in Section 174 Paragraph 4 of the TKG21

¹³⁷<https://www.eworx.at/de/datenschutz>, last visited 2021-12-15

¹³⁸[General Data Protection Regulation, 2016], Art. 13(2) Subpara. f

¹³⁹[General Data Protection Regulation, 2016], Art. 4 Subpara. 4

- the newsletter must contain clear and unambiguous information on its sender
 - the recipient must be provided an authentic address to which they can direct a request to cease such communication
 - the newsletter or any linked websites must not violate Section 6 Paragraph 1 of the ECG
- If the newsletter is sent via a third-party tool, regardless of the type of newsletter:
 - the third party receiving personal data must be disclosed to the data subject at the time of data collection

Special cases

Web beacons:

From a technological standpoint, web beacons ("tracking pixels") are stored on terminal equipment whenever the email containing the web beacon is being downloaded onto the user's computer or smartphone by their email client. Even in case the email is viewed in a web mail client, the pixel would be stored in the user's browser. This means, the use of web beacons is regulated by the Directive on privacy and electronic communications. Web beacons are definitely not strictly necessary and therefore do not fall under the consent exemption provided by Section 165 Paragraph 3 of the TKG21 (or other national transpositions of Article 5(3) of the Directive on privacy and electronic communications). In conclusion, active and informed consent by the recipient is required before an email containing a web beacon may be sent to them.

5.3.5 Mobile tracking

Two survey respondents answered that their companies provided apps via Google Play Store and Apple App Store. This section focuses solely on these distribution channels and their built-in tracking mechanisms as there was no other conclusive evidence of mobile tracking being employed by Austria's Top 500 enterprises.

Use of App/Play Store analytics

App developers distributing their apps on Apple App Store and Google Play Store are provided with aggregated information on a number of KPIs (for more details see section 3.1.1) by the respective store^{140,141}. They are not presented with data identifying a single user. Therefore, Apple or Google are the sole "controllers" under the GDPR for any data gathered by their respective stores and responsible to disclose any collection and processing of personal data to their users.

¹⁴⁰<https://support.google.com/googleplay/android-developer/answer/139628>, last visited 2021-12-16

¹⁴¹<https://help.apple.com/app-store-connect/#/itc623752a8d>, last visited 2021-12-16

5.3.6 Cross-device tracking

Based on the number of cookies found related to Facebook pixel and Facebook promoting cross-device tracking in that regard, the assumption is made that Austria's Top 500 enterprises employ cross-device tracking. As there is no clear evidence of any other cross-device tracking technology other than Facebook pixel, this section deals with Facebook pixel exclusively.

First parties can collect a number of attributes and provide them to Facebook via Facebook pixel¹⁴² to track ad conversions, meaning sales made based on Facebook ads on the respective first-party website, across devices. As Facebook is clearly trying to match all data to specific Facebook users,¹⁴³ all collected data should be considered potentially identifying, especially if the first party adds advanced matching to their Facebook pixel code.¹⁴⁴ In this case, form fields, including first and last name or email as well as physical addresses, are handed over to Facebook. Facebook pixel also uses a cookie to track ad conversions, the respective legal implications are already subsumed under section 5.3.1. Other than that, Facebook is clearly processing the gathered data to enhance their own ad targeting.¹⁴⁵ Therefore, it is assumed that in this case Facebook acts as a "joint controller" under Article 26 of the GDPR.

Facebook provides two websites as examples for opt-out options regarding personalised advertising, which their customers should refer to in their privacy policies. At the time this thesis was written, neither of those websites was fully functional.¹⁴⁶

Whenever data is sent to Facebook, it might be transferred outside the EU. The legal implications of such a transfer is discussed in the following section.

5.3.7 Data transferred outside the EU

The GDPR regulates the transfer of personal data outside the EU in its Chapter V. According to webXrays "Aggregated Tracking Attribution" file, the first tracker which is not located in the U.S. ranks 26th with a presence on 2.48% of all scraped pages. Taking this as well as the results of other web scraping methods into account, this section focuses on data transfer to the U.S..

The European Commission provides a list of adequacy decisions on its website. Countries with adequate level or protection according to the European Commission are, among others, Canada, Japan, Israel and the UK.¹⁴⁷ As mentioned in section 5.2.1, two court

¹⁴²<https://developers.facebook.com/docs/facebook-pixel/>, last visited 2021-12-15

¹⁴³<https://www.facebook.com/business/goals/retargeting>, last visited 2021-12-22

¹⁴⁴facebook.com/business/help/611774685654668?id=1205376682832142, last visited 2021-12-15

¹⁴⁵<https://developers.facebook.com/docs/facebook-pixel/implementation/conversion-tracking/>, last visited 2021-12-22

¹⁴⁶<https://www.facebook.com/legal/terms/business-tools>, last visited 2021-12-15

¹⁴⁷https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en, last visited 2021-12-15

cases involving Maximilian Schrems were the reason for two adequacy decisions being declared invalid: the U.S.-EU Safe Harbor Framework¹⁴⁸ and the EU-U.S. Privacy Shield¹⁴⁹. In conclusion, the only option for data transfer to the U.S. in compliance with the GDPR is to provide appropriate safeguards. Those must ensure the following: *"[...] when personal data are transferred from the Union to controllers, processors or other recipients in third countries or to international organisations, the level of protection of natural persons ensured in the Union by this Regulation should not be undermined, including in cases of onward transfers of personal data from the third country or international organisation to controllers, processors in the same or another third country or international organisation."*¹⁵⁰

Article 46(2) lists options with which appropriate safeguards could be provided, without specific authorisation by the supervisory authority (in Austria the supervisory authority would be the DSB). Most major U.S. companies, like Facebook¹⁵¹ or Google¹⁵², rely on the standard data protection clauses adopted by the Commission in accordance with the examination procedure referred to in Article 93(2)¹⁵³. The latest version of standard contractual clauses (SCC) for international transfers have been made available in June 2021.¹⁵⁴ Before that, a set of SCC adopted under the Directive 95/46/EC existed. Since September 2021, new contracts may only include the new SCC, while existing contracts may still rely on the old ones until December 2022.¹⁵⁵

As mentioned in section 5.2.2 several national data protection authorities, including the Austrian one, as well as the European Data Protection Supervisor decided that Google's use of the SCC for Google Analytics and resulting data transfers to the U.S. does not provide sufficient protection as required by Article 44 of the GDPR.

Controllers are obligated to disclose any intention to transfer personal data to countries outside the EU to the data subject at the time of data collection (for original wording see legal text 5.13). In fig. 5.10 an example of such disclosure can be seen; in this case the company does not rely on the SCC but explicit consent by the data subject under Article 49(1) Subparagraph a.

¹⁴⁸[Schrems (C-362/14), 2015], Para. 98

¹⁴⁹[Facebook Ireland and Schrems (C-311/18), 2020], Para. 201

¹⁵⁰[General Data Protection Regulation, 2016], Rec. 101

¹⁵¹<https://www.facebook.com/help/566994660333381>, last visited 2021-12-15

¹⁵²<https://policies.google.com/privacy/frameworks>, last visited 2021-12-15

¹⁵³[General Data Protection Regulation, 2016], Art. 46(2) Subpara. c

¹⁵⁴https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/standard-contractual-clauses-scc/standard-contractual-clauses-international-transfers_en, last visited 2021-12-15

¹⁵⁵https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/standard-contractual-clauses-scc_en, last visited 2021-12-15

¹⁵⁶[General Data Protection Regulation, 2016], Art.6(1) Subpara. f

¹⁵⁷<https://porr.at/>, last visited 2021-12-13

5. LEGAL ANALYSIS: USER TRACKING IN AUSTRIA AND THE EUROPEAN UNION

Our website uses cookies

Click on "Allow all cookies" to accept all types of cookies or click on "Only allow selected cookies" to allow only those cookies which you may have selected and those which are necessary for the operation and function of the website. You can revoke your decision at any time by changing the cookie settings.

You can find more information here: [Privacy statement](#).

Statistics Marketing

Show details

Required (15)

Technically required cookies are used to enable the technical operation of a website and make it functional for you. The use is based on our legitimate interest to provide a technically flawless website. However, you can generally disable the use of cookies in your browser.

Summary	Creator	Storage time	Domain
BCSF-CS-Cookie	BlueCoat	1 year	port.al
CONSENT	YouTube	2 years	youtube-nocookie.com
CookieConsent	CookieConsent	6 months	port.al
cookieconsent_mode	DataReporter GmbH	12 months	port.al
cookieconsent_status	DataReporter GmbH	12 months	port.al

ensures the functionality and usability of the website and makes it possible to use the features offered by the website.

ensures the functionality and usability of the page and is used to track errors.

ensures the functionality and usability of the page and is used to track errors.

ensures the functionality and usability of the page and is used to track errors.

ensures the functionality and usability of the page and is used to track errors.

Marketing (6)

Marketing cookies come from external advertising companies and are used to collect information about the websites visited by the user. A user gives permission only with your consent and only as long as you have not deactivated the respective cookie. By agreeing to the use of cookies from these service providers, you also expressly consent to the processing of your data in the USA (Art. 49(1)(a) DSGVO).

Summary	Creator	Storage time	Domain
f	Facebook	3 months	.facebook.com
NID	Google	6 months	www.google.com
VISITOR_INFO1_LIVE	YouTube	6 months	youtube.com

registers a unique ID that identifies and recognizes the user. Used for targeted advertising.

registers a unique ID that identifies and recognizes the user. Used for targeted advertising.

registers a unique ID that identifies and recognizes the user. Used for targeted advertising.

OFF

Show details

(a) Example for clear "Accept"-button when using U.S. providers

(b) Example for clear information about the lack of data protection for EU citizens in the U.S.

Figure 5.10: Example: Company disclosing its intention to transfer data to the U.S. and potential risks involved¹⁵⁷

Where personal data relating to a data subject are collected from the data subject, the controller shall, at the time when personal data are obtained, provide the data subject with all of the following information:

[...]

(f) where applicable, the fact that the controller intends to transfer personal data to a third country or international organisation and the existence or absence of an adequacy decision by the Commission, or in the case of transfers referred to in Article 46 or 47, or the second subparagraph of Article 49(1), reference to the appropriate or suitable safeguards and the means by which to obtain a copy of them or where they have been made available.

Legal text 5.13: Article 13(1) Subparagraph f of the General Data Protection Regulation¹⁵⁶

5.3.8 Unregulated tracking methods and privacy risks

The tracking methods employed by tools disclosed by the survey respondents in chapter 3 or detected by the web scraping technologies in chapter 4 are mostly covered by existing law. Methods which are not easily detected, and therefore also not covered in this thesis, potentially pose a greater privacy risk. Even if they were regulated by law, a user would not know they were subjected to it (illegally) and would not be able to take proper action.

Another risk lies in the differences between national transpositions of EU legislation. For example, providers with their place of jurisdiction in Austria have to follow laxer rules based on Section 165 Paragraph 3 of the TKG21 than for example their counterparts in Germany which have to comply to Section 25 Paragraph 1 of the TTDSG. Said differences impact all (EU) users, not only those located in Austria.

Other than that, the way legal requirements are implemented into technology majorly impacts its success in regulating privacy invading practices. Utz et al., among others, studied the implementations of consent notices and published their work in 2019. They found that "*in a privacy-by-default (opt-in) setting, less than 0.1% of visitors allow cookies to be set for all purposes.*"¹⁵⁸ Additionally they state: "*A common motivation to give consent is the assumption that the website cannot be accessed otherwise.*"¹⁵⁸

The number of complaints filed and fines issued under the GDPR is considerably low considering the number of people aware of the legislation and the ease of filing a complaint (relative to other legal actions). In 2019, the legislation was known to 67% of all Europeans (about 343.7 million people), while only 144,376 complaints were filed in the first year after the GDPR came into effect.¹⁵⁹ Based on the website GDPR Enforcement Tracker, which gathers information on fines issued under the GDPR, only 887 complaints have resulted in fines until December 2021.¹⁶⁰ Not all fines are disclosed to the public, the

¹⁵⁸[Utz et al., 2019], p. 974

¹⁵⁹https://ec.europa.eu/info/sites/default/files/infographic-gdpr_in_numbers.pdf, last visited 2021-12-16

¹⁶⁰<https://www.enforcementtracker.com/?insights>, last visited 2021-12-16

number is therefore most likely higher. Considering the results of this thesis, there is no basis to assume these very low numbers may be a result of high GDPR compliance by controllers and/or processors.

Finally, recent research has shown that the GDPR also had adverse effects and actually strengthened the position of large online platforms, like Facebook and Google, for example by reducing their competition.¹⁶¹ This position is also supported by several findings of this thesis, which showed the overwhelming presence of large online platforms as third parties.

¹⁶¹[Geradin et al., 2021], p. 48 ff. & 62 ff.

CHAPTER

6

Summary and future work

6.1 Summary

This thesis contributes to a clearer picture on existing research on utilisation and legal aspects of user tracking methods.

Information was gathered from fragmented research areas into one conclusive chapter on web and mobile tracking. The list of methods provided is non-exhaustive as innovation in the field of user tracking is fast moving and (public) research as well as legislation is often one step behind the current state-of-the-art.

Not all methods identified in chapter 2 were further analysed within this thesis as their use could not conclusively been proven by the methods applied in chapter 3 or chapter 4. The results from these chapters clearly show that Austrian enterprises employ a wide range of user tracking methods for business purposes. Even though the number of responses to the survey conducted for this thesis were limited, the results aided the composition of chapter 5. The additionally conducted web scraping revealed, among others, the utilisation of first- and third-party cookies, other DOM storage methods, tracking via JS files and fingerprinting methods. It also showed that a significant number of user tracking is done through tools provided by third parties, especially major U.S. online platforms, like Google and Facebook.

Based on these findings, the legal analysis performed in chapter 5 revealed that many of the tracking methods in question are already regulated by Austrian or EU law. However, the aforementioned findings also revealed that many companies do not yet apply all of the regulations to their full extent. Additionally, this thesis showed examples of so called dark patterns being employed in order to obtain consent by the highest possible number of users wherever such consent is required by law. This thesis also provides clear evidence that the Austrian transposition of Article 5(3) of the Directive on privacy and electronic communications reduces its applicability to several user tracking methods and

that therefore Austrian enterprises are held to a laxer set of rules than other European companies. In addition, the further analysis of findings in section 4.3 shows that third parties in some cases give false information to the respective first parties, e.g. YouTube's information on their "nocookie" domain (see section 4.3.3), and do not allow for a proper risk assessment. This hinders the lawful utilisation of such tools significantly.

In conclusion, this thesis, while, due to the extent of the field, certainly not exhaustively covering the subject, provides a solid foundation for further research. The following section on future work includes examples of topics in the field not yet covered by this work or upcoming legislation which will impact the findings of this thesis.

6.2 Future work

This thesis would have benefited from a full-scale literature review into web and mobile user tracking methods. Due to the limited resources of a single researcher, this could not be realised. However, without a clear picture on the full scope of state-of-the-art user tracking methods all research conducted based on such fragmented information is inherently incomplete as well.

As was already mentioned in section 3.3, the webXray data could potentially reveal more information on utilised user tracking methods if a more exhaustive analysis were to be performed. It would be especially interesting to deep-dive into the content of certain JS files.

Customer loyalty programs have made use of user tracking methods since their invention. As many of them combine methods not directly connected to web and mobile tracking with data collected using tracking methods employed on the web or mobile phones, those programs are not covered in this thesis. However, one of the biggest fines under the GDPR in Austria issued to a company providing a major Austrian customer loyalty program¹ definitely warrants further research into their employed user tracking methods and the respective legal implications.

Two areas which have not been covered in this thesis because their utilisation by Austrian enterprises was not proven are user tracking methods collecting health and activity data as well as analytics tracking employed on mobile phones. In case of health data, the categorisation as sensitive data under the GDPR means any user tracking concerning such data is regulated by even stricter rules than the processing of other categories of personal data.

The last topics I would like to propose for future work in this field are shared ID, Facebook's Conversion API² and Googles Topics API.³ I expect all of them to provide

¹<https://kurier.at/wirtschaft/joe-bonus-club-soll-2-millionen-euro-datenschutz-strafe-zahlen/401461441>, last visited 2021-12-16

²<https://developers.facebook.com/docs/marketing-api/conversions-api/>, last visited 2021-12-23

³<https://developer.chrome.com/docs/privacy-sandbox/topics/>, last visited 2022-04-22

excellent material to discuss the legal finesse of consent under the GDPR.

From a legal standpoint, there are a few upcoming legislative texts impacting the assessments made in chapter 5. The greatest change would be created by the ePrivacy Regulation (ePR). It was adopted by the European Commission in 2017⁴ and intended to come into effect on 25 May 2018, together with the GDPR. Like the GDPR repealed Directive 95/46/EC, it should have repealed the Directive on privacy and electronic communications.⁵ However, the ePR is still going through ordinary legislative procedure⁶ and is currently being discussed by the Council of the European Union.⁴ The draft was last updated in February 2021.⁷ In its current version, the ePR would complement the GDPR⁸ and protect the *"fundamental rights and freedoms of legal persons in the provision and use of the electronic communications services, and in particular their rights to respect of communications."*⁹ The current estimate is that the ePR will not come into effect until 2025.⁷

On 16 December 2021, the European Parliament adopted the Digital Markets Act (DMA), a regulation defining so-called "gatekeepers" and creating new obligations regarding, among others, personalised advertising and transparency about collected personal data. Shortly before the approval the European Parliament provided the following information: *"The approved text will then become Parliament's mandate for negotiations with EU governments, planned to start under the French presidency of the Council in the first semester of 2022."*¹⁰ A second proposal, the Digital Service Act (DSA), will be voted on by the Internal Market and Consumer Protection Committee in December 2021. This regulation will amend the Directive on electronic commerce. The DSA is assumed to not come into effect until 2024.¹¹

⁴<https://eur-lex.europa.eu/legal-content/EN/HIS/?uri=CELEX%3A52017PC0010>, last visited 2021-12-01

⁵<https://cms.law/en/deu/insight/e-privacy>, last visited 2021-12-01

⁶[Treaty on the Functioning of the European Union, 2012], Art. 294

⁷<https://data.consilium.europa.eu/doc/document/ST-6087-2021-INIT/en/pdf>, last visited 2021-12-01

⁸<https://data.consilium.europa.eu/doc/document/ST-6087-2021-INIT/en/pdf>, Recital 2a, last visited 2021-12-01

⁹<https://data.consilium.europa.eu/doc/document/ST-6087-2021-INIT/en/pdf>, Article 1(1a), last visited 2021-12-01

¹⁰<https://www.europarl.europa.eu/news/en/press-room/20211118IPR17636/digital-markets-act-ending-unfair-practices-of-big-online-platforms>

¹¹<https://www.politico.eu/article/europe-digital-markets-act-dma-digital-services-act-dsa-regulation-platforms-google-amazon-facebook-apple-microsoft/>, last visited 2021-12-17

APPENDIX A

Appendix: Survey

A.1 Lists: Common user tracking tools

A.1.1 Web analytics tools

- Adobe Analytics
- Chartbeat
- Google Analytics
- Mixpanel
- Crazy Egg
- Kissmetrics
- Matomo
- Open Web Analytics
- StatCounter
- Woopra
- Angelfish
- AT Internet: Web Analytics
- AW Stats
- Clicky
- GoSquared
- Hubspot
- Parse.ly
- SEMrush

- SimilarWeb
- Unica NetInsight
- Visual Website Optimizer
- Webalizer
- Yandex Metrica
- Fathom
- Plausible
- Simple Analytics

A.1.2 Fonts

- Google Fonts
- Font Squirrel
- Typekit

A.1.3 Advertisements

Ad servers

- Google Ad Manager
- Broadstreet
- Kevel
- Revive Ad Server
- smart Ad Server

Ad networks

- Amazon Ads
- AdBlade
- Adcash
- AdMaven
- Bing Ads
- Google Ads
- Propel Media
- SmartyAds
- Taboola
- Vibrant Media

A.1.4 Online shops

- BigCommerce
- Shopify
- Wix
- Squarespace
- WooCommerce
- Prestashop
- Square Online
- Volusion
- Big Cartel
- Ecwid
- Magento
- Salesforce Commerce Cloud
- Shift4Shop

A.1.5 Newsletter

- Mailchimp
- GetResponse
- SendInBlue
- CleverReach
- AWeber
- Constant Contact
- ActiveCampaign
- Campaign Monitor
- ConvertKit
- Drip
- MailerLite
- Mailjet
- rapidmail
- Benchmark
- Clever Elements
- Mailify
- Mailingwork
- Moosend

A.1.6 Apps

List of tools

- AdColony
- AdMob
- InMobi
- ironSource
- Audience Network
- Fyber
- Media.net
- MobFox
- Smaato
- StartApp
- Unity Ads

A.2 Online questionnaire

The following answer options were given in questions 22 and 23 of the questionnaire:

- Question 22:
 - 0-9
 - 10-19
 - 20-49
 - 50-249
 - 250+
 - I do not know
 - I cannot answer this question (e.g. due to legal reasons)
- Question 23:
 - Mining and quarrying
 - Manufacturing
 - Electricity, gas, steam and air conditioning supply
 - Water supply; sewerage, waste management and remediation activities
 - Construction
 - Wholesale and retail trade; repair of motor vehicles and motorcycles

- Transportation and storage
- Accommodation and food service activities
- Information and communication
- Financial and insurance activities
- Real estate activities
- Professional, scientific and technical activities
- Administrative and support service activities
- Other service activities
- I do not know
- I cannot answer this question (e.g. due to legal reasons)

A. APPENDIX: SURVEY

userTrackingTechnologies → base 17.11.2021, 15:14

Page 01
Intro

This survey was created as part of my Master thesis at TU Wien, where I am pursuing my studies in "Business Informatics". It was drawn up entirely in English.

No personal data nor identifying business data is collected. All gathered information will be handled with utmost care. To comply to the requirements of my [Netidee](#) grant, gathered data will be released under CC-BY-SA license.

Any questions about the survey or regarding the gathered data can be directed at tanja.travnicek@student.tuwien.ac.at

1. I am not a robot

ten plus zero is

Page 02
Website

User tracking on websites

The following block of questions focuses on your company's website and common user tracking technologies / tools used in this context. Please answer all questions as thoroughly as possible.

2. Does your company have its own website?

Please select one of the following options:

Yes
 No

I do not know
 I cannot answer this question (e.g. due to legal reasons)

1 Active Filter(s)

Filter WB01/F1
If any of the following options is selected: **2**
Then hide the questionnaire page(s) **Cookies, Analytics, Fonts, Ads, Cookies2, Analytics2, Fonts2, Ads2** (otherwise display them)

Page 03

Cookies

3. Does your company's website use any cookies (a small file stored on a user's computer, created by a website server, containing personal information)?

Please select one of the following options:

- Yes
- No

-
- I do not know
 - I cannot answer this question (e.g. due to legal reasons)

1 Active Filter(s)

Filter WB02/F1

If any of the following options is selected: **1**

Then display question/text **WB09** placed later in the questionnaire (otherwise hide)

Page 04

Cookies2

4. Which type of cookies does your company's website use?

Please select one or more of the following options:

- First-party cookies (cookies created by your own web services)
- Third-party cookies (cookies created by third party services, e.g. Google Analytics)

-
- I do not know
 - I cannot answer this question (e.g. due to legal reasons)

A. APPENDIX: SURVEY

Page 05
Analytics

5. Does your company analyse traffic and/or user behaviour on its website (e.g. via Adobe Analytics)?

Please select one of the following options:

- Yes
- No

-
- I do not know
 - I cannot answer this question (e.g. due to legal reasons)

1 Active Filter(s)

Filter WB03/F1

If any of the following options is selected: **1**
Then display question/text **WB04** placed later in the questionnaire (otherwise hide)

6. Which of the following web analytics tools does your company use?

Please select one or more of the following options:

- Adobe Analytics
- Chartbeat
- Google Analytics
- Mixpanel
- Crazy Egg
- Kissmetrics
- Matomo
- Open Web Analytics
- StatCounter
- Woopra
- Angelfish
- AT Internet: Web Analytics
- AW Stats
- Clicky
- GoSquared
- Hubspot
- Parse.ly
- SEMrush
- SimilarWeb
- Unica NetInsight
- Visual Website Optimizer
- Webalizer
- Yandex Metrica

Other (please specify):

-
- I do not know
 - I cannot answer this question (e.g. due to legal reasons)

A. APPENDIX: SURVEY

Page 07

Fonts

7. Does your company's website include any fonts from font directories (e.g. Google Fonts)?

Please select one of the following options:

- Yes
- No

-
- I do not know
 - I cannot answer this question (e.g. due to legal reasons)

1 Active Filter(s)

Filter WB05/F1

If any of the following options is selected: **1**

Then display question/text **WB06** placed later in the questionnaire (otherwise hide)

Page 08

Fonts2

8. Which of the following font directories does your company use?

Please select one or more of the following options:

- Google Fonts
- Font Squirrel
- Typekit
- Other (please specify):

-
- I do not know
 - I cannot answer this question (e.g. due to legal reasons)

Page 09

Ads

9. Does your company's website display any third party advertisements?

Please select one of the following options:

- Yes
- No

- I do not know
- I cannot answer this question (e.g. due to legal reasons)

1 Active Filter(s)

Filter WB07/F1

If any of the following options is selected: **1**

Then display question/text **WB08** placed later in the questionnaire (otherwise hide)

Page 10

Ads2

10. Which of the following ad networks and/or ad servers does your company use?

Please select one or more of the following options:

- Amazon Ads
- AdBlade
- Adcash
- AdMaven
- Bing Ads
- Google Ads
- Propel Media
- SmartyAds
- Taboola
- VerizonMedia
- Vibrant Media
- Google Ad Manager
- Broadstreet
- Kevel
- Revive Ad Server
- smart Ad Server

Other (please specify):

-
- I do not know
 - I cannot answer this question (e.g. due to legal reasons)

User tracking in online shops

The following block of questions focuses on your company's online shop and common user tracking technologies / tools used in this context. Please answer all questions as thoroughly as possible.

11. Does your company have its own online shop?

Please select one of the following options:

- Yes
- No

-
- I do not know
 - I cannot answer this question (e.g. due to legal reasons)

1 Active Filter(s)

Filter OS01/F1

If any of the following options is selected: **2**

Then hide the questionnaire page(s) **OSTools**, **OSTools2** (otherwise display them)

12. Does your company use any existing web shop technologies offered by e-commerce platforms (e.g. BigCommerce) ?

Please select one of the following options:

- Yes
- No

-
- I do not know
 - I cannot answer this question (e.g. due to legal reasons)

1 Active Filter(s)

Filter OS02/F1

If any of the following options is selected: **1**

Then display question/text **OS03** placed later in the questionnaire (otherwise hide)

13. Which of the following eCommerce platforms does your company use?

Please select one or more of the following options:

- BigCommerce
- Shopify
- Wix
- Squarespace
- WooCommerce
- Prestashop
- Square Online
- Volusion
- Big Cartel
- Ecwid
- Magento
- Salesforce Commerce Cloud
- Shift4Shop

Other (please specify):

-
- I do not know
 - I cannot answer this question (e.g. due to legal reasons)

User tracking via newsletters

The following block of questions focuses on your company's newsletter and common user tracking technologies / tools used in this context. Please answer all questions as thoroughly as possible.

14. Does your company send out newsletters?

Please select one of the following options:

- Yes
- No

-
- I do not know
 - I cannot answer this question (e.g. due to legal reasons)

1 Active Filter(s)

Filter NL01/F1

If any of the following options is selected: **2**

Then hide the questionnaire page(s) **NLTools, NLTools2** (otherwise display them)

15. Does your company use a third-party newsletter tool (e.g. Mailchimp)?

Please select one of the following options:

- Yes
- No

-
- I do not know
 - I cannot answer this question (e.g. due to legal reasons)

1 Active Filter(s)

Filter NL02/F1

If any of the following options is selected: **1**

Then display question/text **NL03** placed later in the questionnaire (otherwise hide)

16. Which of the following newsletter tools does your company use?

Please select one or more of the following options:

- Mailchimp
- GetResponse
- SendInBlue
- CleverReach
- AWeber
- Constant Contact
- ActiveCampaign
- Campaign Monitor
- ConvertKit
- Drip
- MailerLite
- Mailjet
- rapidmail
- Benchmark
- Clever Elements
- Mailify
- Mailingwork
- Moosend

Other (please specify):

- I do not know
- I cannot answer this question (e.g. due to legal reasons)

User tracking in mobile applications

The following block of questions focuses on your company's applications for mobile devices and common user tracking technologies / tools used in this context. Please answer all questions as thoroughly as possible.

17. Does your company provide any applications for mobile devices?

Please select one of the following options:

- Yes
- No

- I do not know
- I cannot answer this question (e.g. due to legal reasons)

1 Active Filter(s)

Filter MA01/F1

If any of the following options is selected: **2**
Then hide the questionnaire page(s) **AppStore, IAAds, IAAds2** (otherwise display them)

18. Does your company use Google Play Store and/or Apple App Store for application distribution?

Please select one of the following options:

- Yes, both
- Yes, only Apple App Store
- Yes, only Google Play Store
- No

- I do not know
- I cannot answer this question (e.g. due to legal reasons)

Page 19

IAAds

19. Do your company's mobile applications display any third party advertisements?

Please select one of the following options:

- Yes
- No

- I do not know
- I cannot answer this question (e.g. due to legal reasons)

1 Active Filter(s)

Filter MA03/F1

If any of the following options is selected: **1**

Then display question/text **MA04** placed later in the questionnaire (otherwise hide)

Page 20

IAAds2

20. Which of the following mobile ad networks does your company use?

Please select one or more of the following options:

- AdColony
- AdMob
- InMobi
- ironSource
- Audience Network
- Fyber
- Media.net
- MobFox
- Smaato
- StartApp
- Unity Ads

Other (please specify):

-
- I do not know
 - I cannot answer this question (e.g. due to legal reasons)

Further user tracking

The following question tries to reduce any blind spots in my Master thesis. Please answer it as thoroughly as possible.

21. Does your company use any user tracking tools not mentioned in this survey?

Please select one of the following options:

Yes (please specify):

No

I do not know

I cannot answer this question (e.g. due to legal reasons)

Page 22
Company

Statistical values of your company

The following questions are used for statistical evaluation of the companies providing the collected data only. All answers still remain completely anonymous.

22. How many employees does your company have?

Please select one of the following options:

[Please choose] ▾

23. What is your company's main economic activity?

Please select one of the following options:

[Please choose] ▾

24. Where does your company operate?

Please select one or more of the following options:

- Austria
- Other countries within the European Union
- Countries outside the European Union

-
- I do not know
 - I cannot answer this question (e.g. due to legal reasons)

Last Page

Thank you for your participation

Your answers have been recorded. You can close this page and/or this browser now.

Any questions about the survey or regarding the gathered data can be directed at
tanja.travnicek@student.tuwien.ac.at

List of Figures

2.1	Visualisation of the search process for item (b) of the search phase	6
2.2	Tagged publications grouped by publication year	8
2.3	Tagged publications grouped by content	9
2.4	History of web tracking by Bujlow et al. from 1990 to 2015	11
2.5	Rise in tracking domains from 1996 to 2016	12
2.6	Figure 6 from Lerner et al.'s "Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study of Web Tracking from 1996 to 2016" .	14
2.7	Top 10 attributed collected by fingerprinters, from Al-Fannah et al.'s 2018 study	17
2.8	JavaScript-based font probing scripts on homepages of Top 1 Million Alexa sites from Acar et al.'s paper	18
2.9	The principle of email tracking as described by Bender et al. 2016 and Fabian et al. 2021	21
2.10	Heuristics to decide on app installation (Kulyk et al., 2016)	23
2.11	High-level statistics for illustrative crawls under the two third-party cookie settings by Acar et al.	29
2.12	Relative percentage, based on the state as of December 2018, of the number of publishers of popular and less popular trackers and CNAME-based trackers from Dimova et al.'s paper	31
3.1	Flowchart of the questionnaire	42
4.1	The Top 24 first-party cookies, which were found most on the websites of Austria's Top 500 enterprises on page load	53
4.2	The Top 25 first-party cookies, which were found most on the websites of Austria's Top 500 enterprises when browsing with Consent-O-Matic extension	56
4.3	The Top 25 first-party cookies, which were found most on the websites of Austria's Top 500 enterprises when browsing with "I don't care about cookies" extension	57
4.4	Comparison of first 20 URLs sorted descending by the number of cookies present with and without extensions	59
4.5	Comparison of first 20 URLs sorted descending by the number of script sources present with and without extensions	61

4.6	First 20 URLs sorted descending by the number of third-party cookies sources found by webXray	62
4.7	Top 15 third-party scripts found by webXray, sorted by percentage of presence on websites provided by Austria's Top 500 enterprises	63
4.8	YouTube's dialog for video embed code generation	65
4.9	First 20 URLs sorted descending by the number of DOM storage entries found by webXray	66
4.10	Top 10 third-party scripts found by webXray, sorted by their presence in absolute numbers on websites provided by Austria's Top 500 enterprises .	66
5.1	Example: Consent notice with insufficient information and no proper consent option ¹	78
5.2	Example: Consent notice with preticked box ²	80
5.3	Example: Insufficient cookie information ³	81
5.4	Example: Well-defined cookie information ⁴	81
5.5	Example: Rejection option on the first layer of a cookie consent pop-up ⁵ .	83
5.6	Example: Utilisation of a dark pattern ⁶	83
5.7	Example: False cookie categorisation ⁷	84
5.8	Example: Missing information on strictly necessary cookie ⁸	85
5.9	Example: Information on HTML5 localStorage objects ⁹	89
5.10	Example: Company disclosing its intention to transfer data to the U.S. and potential risks involved ¹⁰	100

List of Listings

4.1	Python script to scrape for cookies	50
4.2	Python script to scrape for script sources	51
4.3	Code snippet to use a Chrome extension with Selenium and ChromeDriver	52

List of Legal texts

5.1 Article 7 of the Charter of Fundamental Rights of the European Union ¹¹	75
5.2 Article 8 of the Charter of Fundamental Rights of the European Union ¹²	75
5.3 Article 5(3) of Directive on privacy and electronic communications ¹³	75
5.4 Section 165 Paragraph 3 of the Telekommunikationsgesetz 2021 (German original) ¹⁴	76
5.5 Section 3 Subparagraph 1 of E-Commerce Gesetz in its official English version ¹⁵	77
5.6 Article 5(1) Paragraph e of the General Data Protection Regulation ¹⁶	78
5.7 Article 4 Subparagrah 1 of the General Data Protection Regulation ¹⁷	79
5.8 Article 4 Subparagraph 11 of the General Data Protection Regulation ¹⁸	79
5.9 Article 2 Subparagraph f of the Directive on electronic commerce ¹⁹	93
5.10 Section 3 Subparagraph 6 of the E-Commerce Gesetz in its official english translation ²⁰	94
5.11 Section 174 Paragraph 4 of the Telekommunikationsgesetz 2021 ²¹	95
5.12 Article 4 Subparagraph 4 of the General Data Protection Regulation ²²	96
5.13 Article 13(1) Subparagraph f of the General Data Protection Regulation ²³	101

Acronyms

A record Address record. 30

ABGB Allgemeines Bürgerliches Gesetzbuch. 73

API application programming interface. 14, 29, 104

CFR Charter of Fundamental Rights of the European Union. 1, 69, 73–75, 79, 131

CJEU Court of Justice of the European Union. 68, 70–72, 74

CMP consent management platform. 48, 55, 57, 58, 60, 72, 82, 83, 89, 92

CNAME Canonical Name. 30, 60

CRX Chrome extension file. 51

CSS Cascading Style Sheets. 37, 65

CSV comma-separated values. 49, 52, 57

DMA Digital Markets Act. 105

DNS Domain Name System. 10, 30, 54

DOM Document Object Model. ix, xi, 64, 66, 103, 128

DoS Denial-of-Service. 86

DPC Data Protection Commissioner. 69–72

DSA Digital Service Act. 105

DSB Datenschutzbehörde. 74, 99

DSG Datenschutzgesetz. 73, 75

DSG 2000 Datenschutzgesetz 2000. 73

ECG E-Commerce Gesetz. 73, 76, 77, 93, 94, 96, 97, 131

ECJ European Court of Justice. 11, 70–72, 74, 79, 85, 86

ePR ePrivacy Regulation. 105

EU European Union. ix, xi, 1, 45, 67–74, 84, 85, 90, 92, 93, 98–101, 103, 105

FCC Federal Communications Commission. 20

FLoC Federated Learning of Cohorts. 29

FTC Federal Trade Commission. 26, 27

GDPR General Data Protection Regulation. 1, 25, 41, 69–75, 77–80, 82, 86–88, 91–94, 96–99, 101, 102, 104, 105, 131

HTML HyperText Markup Language. 49

HTTP Hypertext Transfer Protocol. 10, 11, 13–16, 19, 30, 65

HTTPS Hypertext Transfer Protocol Secure. 30

ID identifier. 13–17, 19, 22, 24, 25, 28–30, 40, 60, 104

IP Internet Protocol. 17–19, 30, 37, 71, 91, 92, 95, 96

ISP Internet Service Provider. 16, 19

JS JavaScript. ix, xi, 16–21, 47, 49, 55, 60, 63, 65, 103, 104, 127

JSON JavaScript Object Notation. 65

KPI key performance indicator. 36, 97

PII personally identifiable information. 24, 25

RIS Rechtsinformationssystem des Bundes. 68, 73

SCC standard contractual clauses. 72, 99

TKG21 Telekommunikationsgesetz 2021. 73, 75–80, 82, 83, 85, 87, 88, 90, 93–97, 101, 131

TTDSG Telekommunikation-Telemedien-Datenschutz-Gesetz. 82, 85, 101

U.S. United States (of America). ix, xi, 1, 70–72, 92, 98–100, 103, 128

UK United Kingdom. 98

URL Uniform Resource Locator. 27, 49, 52, 55, 57–62, 66, 127, 128

UUID universally unique identifier. 25, 30, 55

Bibliography

- [Acar et al., 2014] Acar, G., Eubank, C., Englehardt, S., Juarez, M., Narayanan, A., and Diaz, C. (2014). 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. Association for Computing Machinery.
- [Acar et al., 2013] Acar, G., Juarez, M., Nikiforakis, N., Diaz, C., Gürses, S., Piessens, F., and Preneel, B. (2013). Fpdetective: Dusting the web for fingerprinters. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, CCS '13, page 1129–1140, New York, NY, USA. Association for Computing Machinery.
- [Al-Fannah et al., 2018] Al-Fannah, N. M., Li, W., and Mitchell, C. J. (2018). Beyond cookie monster amnesia: Real world persistent online tracking. In Chen, L., Manulis, M., and Schneider, S., editors, *Information Security*, pages 481–501, Cham. Springer International Publishing.
- [Al-Fannah and Mitchell, 2020] Al-Fannah, N. M. and Mitchell, C. (2020). Too little too late: Can we control browser fingerprinting? *Journal of Intellectual Capital*, 21:165–180.
- [Allgemeines bürgerliches Gesetzbuch, JGS 946/1811 idF I 175/2021, 1811] Allgemeines bürgerliches Gesetzbuch, JGS 946/1811 idF I 175/2021 (1811). Allgemeines bürgerliches Gesetzbuch für die gesammten deutschen Erbländer der Oesterreichischen Monarchie.
- [Application - Breyer (C-582/14), 2015] Application - Breyer (C-582/14) (2015). Case C-582/14: Request for a preliminary ruling from the Bundesgerichtshof (Germany) lodged on 17 December 2014 — Patrick Breyer v Bundesrepublik Deutschland. *Office Journal C* 89, page 4–5.
- [Article 29 Data Protection Working Party, 2014] Article 29 Data Protection Working Party (2014). Opinion 9/2014 on the application of Directive 2002/58/EC to device fingerprinting.

- [Ayenson et al., 2011] Ayenson, M. D., Wambach, D. J., Soltani, A., Good, N., and Hoofnagle, C. J. (2011). Flash cookies and privacy II: Now with HTML5 and ETag respawning. *Available at SSRN 1898390*.
- [Bekavac and Garbin Praničević, 2015] Bekavac, I. and Garbin Praničević, D. (2015). Web analytics tools and web metrics tools: An overview and comparative analysis. *Croatian Operational Research Review*, 6(2):373–386.
- [Belloro and Mylonas, 2018] Belloro, S. and Mylonas, A. (2018). I know what you did last summer: New persistent tracking mechanisms in the wild. *IEEE Access*, 6:52779–52792.
- [Bender et al., 2016] Bender, B., Fabian, B., Lessmann, S., and Haupt, J. (2016). E-mail tracking: Status quo and novel countermeasures. In *Thirty Seventh International Conference on Information Systems, Dublin 2016*.
- [Biemer, 2010] Biemer, P. P. (2010). Total Survey Error: Design, Implementation, and Evaluation. *Public Opinion Quarterly*, 74(5):817–848.
- [Binns et al., 2018] Binns, R., Lyngs, U., Van Kleek, M., Zhao, J., Libert, T., and Shadbolt, N. (2018). Third party tracking in the mobile ecosystem. In *Proceedings of the 10th ACM Conference on Web Science*, pages 23–31.
- [Boda et al., 2012] Boda, K., Földes, Á. M., Gulyás, G. G., and Imre, S. (2012). User tracking on the web via cross-browser fingerprinting. In Laud, P., editor, *Information Security Technology for Applications*, pages 31–46, Berlin, Heidelberg. Springer Berlin Heidelberg.
- [Bouguettaya and Eltoweissy, 2003] Bouguettaya, A. and Eltoweissy, M. (2003). Privacy on the web: facts, challenges, and solutions. *IEEE Security Privacy*, 1(6):40–49.
- [Boutet and Gambs, 2019] Boutet, A. and Gambs, S. (2019). Inspect what your location history reveals about you: Raising user awareness on privacy threats associated with disclosing his location data. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM ’19*, page 2861–2864, New York, NY, USA. Association for Computing Machinery.
- [Brace, 2008] Brace, I. (2008). *Questionnaire design: How to plan, structure and write survey material for effective market research*. Kogan Page Publishers.
- [Bujlow et al., 2015] Bujlow, T., Carela-Español, V., Solé-Pareta, J., and Barlet-Ros, P. (2015). Web tracking: Mechanisms, implications, and defenses. *arXiv preprint arXiv:1507.07872*.
- [Burchell and Marsh, 1992] Burchell, B. and Marsh, C. (1992). The effect of questionnaire length on survey response. *Quality and quantity*, 26(3):233–244.

- [Charter of Fundamental Rights of the European Union, 2012] Charter of Fundamental Rights of the European Union (2012). Charter of Fundamental Rights of the European Union. *Office Journal C* 326, page 391–407.
- [Christl et al., 2017] Christl, W., Kopp, K., and Riechert, P. U. (2017). Corporate surveillance in everyday life. *Cracked Labs*, page 6.
- [Christl and Spiekermann, 2016] Christl, W. and Spiekermann, S. (2016). *Networks of control*. Facultas Verlags- und Buchhandels AG.
- [Cohen-Almagor, 2013] Cohen-Almagor, R. (2013). Internet history. In *Moral, ethical, and social dilemmas in the age of technology: Theories and practice*, pages 19–39. IGI Global.
- [Datenschutzgesetz 2000, BGBI. I 165/1999 idF I 120/2017, 1999] Datenschutzgesetz 2000, BGBI. I 165/1999 idF I 120/2017 (1999). Bundesgesetz über den Schutz personenbezogener Daten (Datenschutzgesetz 2000 – DSG 2000).
- [Datenschutzgesetz, BGBI. I 165/1999 idF I 14/2019, 1999] Datenschutzgesetz, BGBI. I 165/1999 idF I 14/2019 (1999). Bundesgesetz zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten (Datenschutzgesetz – DSG).
- [Debusseré, 2005] Debusseré, F. (2005). The EU E-Privacy Directive: A Monstrous Attempt to Starve the Cookie Monster? *International Journal of Law and Information Technology*, 13(1):70–97.
- [Dimova et al., 2021] Dimova, Y., Acar, G., Olejnik, L., Joosen, W., and Van Goethem, T. (2021). The cname of the game: Large-scale analysis of dns-based tracking evasion. *arXiv preprint arXiv:2102.09301*.
- [Directive 2006/24/EC, 2006] Directive 2006/24/EC (2006). Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC. *Office Journal L* 105, page 54–63.
- [Directive 2009/136/EC, 2009] Directive 2009/136/EC (2009). Directive 2009/136/EC of the European Parliament and of the Council of 25 November 2009 amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services, Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector and Regulation (EC) no 2006/2004 on cooperation between national authorities responsible for the enforcement of consumer protection laws. *Office Journal L* 337, pages 11–36.
- [Directive 95/46/EC, 1995] Directive 95/46/EC (1995). Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Office Journal L* 281, page 31–50.

[Directive (EU) 2018/1972, 2018] Directive (EU) 2018/1972 (2018). Directive (EU) 2018/1972 of the European Parliament and of the Council of 11 December 2018 establishing the European Electronic Communications Code (Recast). *Office Journal L 321*, page 36–214.

[Directive on electronic commerce, 2000] Directive on electronic commerce (2000). Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce'). *Office Journal L 178*, pages 1–16.

[Directive on privacy and electronic communications, 2002] Directive on privacy and electronic communications (2002). Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications). *Office Journal L 201*, pages 37–47.

[E-Commerce-Gesetz, BGBI. I 152/2001 idF I 148/2020, 2001] E-Commerce-Gesetz, BGBI. I 152/2001 idF I 148/2020 (2001). Bundesgesetz, mit dem bestimmte rechtliche Aspekte des elektronischen Geschäfts- und Rechtsverkehrs geregelt werden (E-Commerce-Gesetz – ECG).

[Eckersley, 2010] Eckersley, P. (2010). How unique is your web browser? In Atallah, M. J. and Hopper, N. J., editors, *Privacy Enhancing Technologies*, pages 1–18, Berlin, Heidelberg. Springer Berlin Heidelberg.

[Enck et al., 2010] Enck, W., Gilbert, P., Chun, B.-G., Cox, L. P., Jung, J., McDaniel, P., and Sheth, A. N. (2010). Taintdroid: An information-flow tracking system for realtime privacy monitoring on smartphones. In *9th USENIX Symposium on Operating Systems Design and Implementation (OSDI 10)*, Vancouver, BC. USENIX Association.

[Englehardt et al., 2018] Englehardt, S., Han, J., and Narayanan, A. (2018). I never signed up for this! Privacy implications of email tracking. *Proceedings on Privacy Enhancing Technologies*, 2018.

[Ermakova et al., 2018] Ermakova, T., Fabian, B., Bender, B., and Klimek, K. (2018). Web tracking - a literature review on the state of research. In *Proceedings of the 51st Hawaii International Conference on System Sciences*, pages 4732–4741.

[Esteve, 2017] Esteve, A. (2017). The business of personal data: Google, Facebook, and privacy issues in the EU and the USA. *International Data Privacy Law*, 7(1):36–47.

[Eubank et al., 2013] Eubank, C., Melara, M., Perez-Botero, D., and Narayanan, A. (2013). Shining the floodlights on mobile web tracking — a privacy survey.

[Fabian et al., 2021] Fabian, B., Bender, B., Hesseldieck, B., Haupt, J., and Lessmann, S. (2021). Enterprise-grade protection against e-mail tracking. *Information Systems*, 97:101702.

[Facebook Ireland and Schrems (C-311/18), 2020] Facebook Ireland and Schrems (C-311/18) (2020). C-311/18: Judgment of the Court (Grand Chamber) of 16 July 2020 (request for a preliminary ruling from the High Court (Ireland) — Ireland) — Data Protection Commissioner v Facebook Ireland Ltd, Maximillian Schrems (Reference for a preliminary ruling — Protection of individuals with regard to the processing of personal data — Charter of Fundamental Rights of the European Union — Articles 7, 8 and 47 — Regulation (EU) 2016/679 — Article 2(2) — Scope — Transfers of personal data to third countries for commercial purposes — Article 45 — Commission adequacy decision — Article 46 — Transfers subject to appropriate safeguards — Article 58 — Powers of the supervisory authorities — Processing of the data transferred by the public authorities of a third country for national security purposes — Assessment of the adequacy of the level of protection in the third country — Decision 2010/87/EU — Protective standard clauses on the transfer of personal data to third countries — Suitable safeguards provided by the data controller — Validity — Implementing Decision (EU) 2016/1250 — Adequacy of the protection provided by the EU-US Privacy Shield — Validity — Complaint by a natural person whose data was transferred from the European Union to the United States). *Office Journal C* 297, pages 4–5.

[Fashion ID (C-40/17), 2019] Fashion ID (C-40/17) (2019). Case C-40/17: Judgment of the Court (Second Chamber) of 29 July 2019 (request for a preliminary ruling from the Oberlandesgericht Düsseldorf — Germany) — Fashion ID GmbH & Co. KG v Verbraucherzentrale NRW eV (Reference for a preliminary ruling — Protection of individuals with regard to the processing of personal data — Directive 95/46/EC — Article 2(d) — Notion of ‘controller’ — Operator of a website who has embedded on that website a social plugin that allows the personal data of a visitor to that website to be transferred to the provider of that plugin — Article 7(f) — Lawfulness of data processing — Taking into account of the interest of the operator of the website or of that of the provider of the social plugin — Articles 2(h) and 7(a) — Consent of the data subject — Article 10 — Informing the data subject — National legislation allowing consumer-protection associations to bring or defend legal proceedings). *Office Journal C* 319, page 2–3.

[Fink, 2010] Fink, R. (2010). Web fonts at the crossing. *A List Apart*, 307.

[Galesic and Bosnjak, 2009] Galesic, M. and Bosnjak, M. (2009). Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. *Public Opinion Quarterly*, 73(2):349–360.

[General Data Protection Regulation, 2016] General Data Protection Regulation (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Office Journal L* 119, pages 1–88.

- [Geradin et al., 2021] Geradin, D., Karanikoti, T., and Katsifis, D. (2021). Gdpr myopia: how a well-intended regulation ended up favouring large online platforms - the case of ad tech. *European Competition Journal*, 17(1):47–92.
- [Greenhalgh et al., 2005] Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., Kyriakidou, O., and Peacock, R. (2005). Storylines of research in diffusion of innovation: A meta-narrative approach to systematic review. *Social science & medicine*, 61(2):417–430.
- [Hill, 1998] Hill, R. (1998). What sample size is “enough” in internet survey research. *Interpersonal Computing and Technology: An electronic journal for the 21st century*, 6(3-4):1–12.
- [Hu and Sastry, 2020] Hu, X. and Sastry, N. (2020). What a tangled web we weave: Understanding the interconnectedness of the third party cookie ecosystem. In *12th ACM Conference on Web Science*, WebSci ’20, page 76–85, New York, NY, USA. Association for Computing Machinery.
- [Judgment - Breyer (C-582/14), 2016] Judgment - Breyer (C-582/14) (2016). C-582/14: Judgment of the Court (Second Chamber) of 19 October 2016 (request for a preliminary ruling from the Bundesgerichtshof) Patrick Breyer v Bundesrepublik Deutschland (Reference for a preliminary ruling — Processing of personal data — Directive 95/46/EC — Article 2(a) — Article 7(f) — Definition of ‘personal data’ — Internet protocol addresses — Storage of data by an online media services provider — National legislation not permitting the legitimate interest pursued by the controller to be taken into account). *Office Journal C* 475, page 3.
- [Kulyk et al., 2016] Kulyk, O., Gerber, P., Hanafi, M., Berens, B., Renaud, K., and Volkamer, M. (2016). Encouraging privacy-aware smartphone app installation: What would the technically-adept do. In *NDSS Workshop on Usable Security (USEC)*.
- [Lerner et al., 2016] Lerner, A., Simpson, A. K., Kohno, T., and Roesner, F. (2016). Internet jones and the raiders of the lost trackers: An archaeological study of web tracking from 1996 to 2016. In *25th {USENIX} Security Symposium ({USENIX} Security 16)*.
- [Libert, 2015] Libert, T. (2015). Exposing the invisible web: An analysis of third-party HTTP requests on 1 million websites. *International Journal of Communication*, 9(0):3544–3561.
- [Lin et al., 2014] Lin, J., Liu, B., Sadeh, N., and Hong, J. I. (2014). Modeling users’ mobile app privacy preferences: Restoring usability in a sea of permission settings. In *10th Symposium On Usable Privacy and Security (SOUPS 2014)*, pages 199–212, Menlo Park, CA. USENIX Association.

- [Lockhart et al., 2012] Lockhart, J. W., Pulickal, T., and Weiss, G. M. (2012). Applications of mobile activity recognition. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 1054–1058.
- [Maletic and Marcus, 2010] Maletic, J. I. and Marcus, A. (2010). Data cleansing: A prelude to knowledge discovery. In Maimon, O. and Rokach, L., editors, *Data Mining and Knowledge Discovery Handbook*, pages 19–32. Springer US, Boston, MA.
- [Malgieri and Custers, 2018] Malgieri, G. and Custers, B. (2018). Pricing privacy – the right to know the value of your personal data. *Computer Law & Security Review*, 34(2):289–303.
- [Maximilian Schrems v Facebook Ireland Ltd (C-446/21), 2021] Maximilian Schrems v Facebook Ireland Ltd (C-446/21) (2021). Case C-446/21: Request for a preliminary ruling from the Oberster Gerichtshof (Austria) lodged on 20 July 2021 — Maximilian Schrems v Facebook Ireland Ltd. *Office Journal C 422*, page 5–6.
- [Mayer, 2009] Mayer, J. R. (2009). "Any person... a pamphleteer": Internet anonymity in the age of web 2.0. In *Princeton School of Public and International Affairs, 1929-2021*. Princeton Research Data Service.
- [Mayer and Mitchell, 2012] Mayer, J. R. and Mitchell, J. C. (2012). Third-party web tracking: Policy and technology. In *2012 IEEE Symposium on Security and Privacy*, pages 413–427.
- [McKay et al., 2019] McKay, D., Chua, Y. J., Chang, S., Dreyfus, S., Whitty, M., Paterson, J. M., Zhan, P., Hanley, G., and Clausen, A. (2019). State of the art in data tracking technology.
- [Meadows, 2013] Meadows, A. (2013). *ASP. NET MVC 4 mobile app development*. Packt Publishing Ltd.
- [Meissel, 2017] Meissel, F.-S. (2017). Allgemeines Bürgerliches Gesetzbuch. In *Staatslexikon der Görres-Gesellschaft*, volume 8, pages 112–115. Görres-Gesellschaft / Verlag Herder, 1 edition.
- [Mellet and Beauvisage, 2019] Mellet, K. and Beauvisage, T. (2019). Cookie monsters. Anatomy of a digital market infrastructure. *Consumption Markets & Culture*, 23:1–20.
- [Menéndez, 2002] Menéndez, A. J. (2002). Chartering europe: Legal status and policy implications of the charter of fundamental rights of the european union. *JCMS: Journal of Common Market Studies*, 40(3):471–490.
- [Mishra et al., 2020] Mishra, V., Laperdrix, P., Vastel, A., Rudametkin, W., Rouvoy, R., and Lopatka, M. (2020). Don't count me out: On the relevance of IP address in the tracking ecosystem. In *Proceedings of The Web Conference 2020*, WWW '20, page 808–815, New York, NY, USA. Association for Computing Machinery.

- [Mitchell, 2018] Mitchell, R. (2018). *Web scraping with Python: Collecting more data from the modern web*. O'Reilly Media, Inc.
- [Montulli, 1995] Montulli, L. (U.S. patent 5 774 670 A, 1995). Persistent client state in a hypertext transfer protocol based client-server system.
- [Mowery and Shacham, 2012] Mowery, K. and Shacham, H. (2012). Pixel perfect: Fingerprinting canvas in HTML5. In Fredrikson, M., editor, *Proceedings of W2SP 2012*. IEEE Computer Society.
- [Nikiforakis et al., 2013] Nikiforakis, N., Kapravelos, A., Joosen, W., Kruegel, C., Piessens, F., and Vigna, G. (2013). Cookieless monster: Exploring the ecosystem of web-based device fingerprinting. In *2013 IEEE Symposium on Security and Privacy*, pages 541–555.
- [Norwegian Consumer Council, 2018] Norwegian Consumer Council (2018). Deceived by design, how tech companies use dark patterns to discourage us from exercising our rights to privacy. *Norwegian Consumer Council Report*.
- [Pachilakis et al., 2019] Pachilakis, M., Papadopoulos, P., Markatos, E. P., and Kourtellis, N. (2019). No more chasing waterfalls: A measurement study of the header bidding ad-ecosystem. In *Proceedings of the Internet Measurement Conference*, IMC '19, page 280–293, New York, NY, USA. Association for Computing Machinery.
- [Peacock, 2014] Peacock, S. E. (2014). How web tracking changes user agency in the age of big data: The used user. *Big Data & Society*, 1(2):2053951714564228.
- [Planet49 (C-673/17), 2019] Planet49 (C-673/17) (2019). C-673/17: Judgment of the Court (Grand Chamber) of 1 October 2019 (request for a preliminary ruling from the Bundesgerichtshof — Germany) — Bundesverband der Verbraucherzentralen und Verbraucherverbände — Verbraucherzentrale Bundesverband eV v Planet49 GmbH (Reference for a preliminary ruling — Directive 95/46/EC — Directive 2002/58/EC — Regulation (EU) 2016/679 — Processing of personal data and protection of privacy in the electronic communications sector — Cookies — Concept of consent of the data subject — Declaration of consent by means of a pre-ticked checkbox). *Office Journal C 413*, page 4.
- [Regulation (EU) 2017/2394, 2017] Regulation (EU) 2017/2394 (2017). Regulation (EU) 2017/2394 of the European Parliament and of the Council of 12 December 2017 on cooperation between national authorities responsible for the enforcement of consumer protection laws and repealing Regulation (EC) No 2006/2004. *Office Journal L 345*, page 1–26.
- [Roscoe, 1975] Roscoe, J. T. (1975). *Fundamental research statistics for the behavioral sciences*, volume 2. Holt, Rinehart and Winston.

- [Sanchez-Rola et al., 2020] Sanchez-Rola, I., Balzarotti, D., and Santos, I. (2020). Cookies from the past: Timing server-side request processing code for history sniffing. *Digital Threats: Research and Practice*, 1(4).
- [Sanchez-Rola et al., 2016] Sanchez-Rola, I., Ugarte-Pedrero, X., Santos, I., and Bringas, P. G. (2016). The web is watching you: A comprehensive review of web-tracking techniques and countermeasures. *Logic Journal of the IGPL*, 25(1):18–29.
- [Scarlet Extended (C-311/18), 2011] Scarlet Extended (C-311/18) (2011). Case C-70/10: Judgment of the Court (Third Chamber) of 24 November 2011 (reference for a preliminary ruling from the Cour d'appel de Bruxelles (Belgium)) — Scarlet Extended SA v Société belge des auteurs, compositeurs et éditeurs SCRL (SABAM) (Information society — Copyright — Internet — ‘Peer-to-peer’ software — Internet service providers — Installation of a system for filtering electronic communications in order to prevent file sharing which infringes copyright — No general obligation to monitor information transmitted). *Office Journal C* 25, page 6–7.
- [Schrems (C-362/14), 2015] Schrems (C-362/14) (2015). C-362/14: Judgment of the Court (Grand Chamber) of 6 October 2015 (request for a preliminary ruling from the High Court (Ireland)) — Maximillian Schrems v Data Protection Commissioner (Reference for a preliminary ruling — Personal data — Protection of individuals with regard to the processing of such data — Charter of Fundamental Rights of the European Union — Articles 7, 8 and 47 — Directive 95/46/EC — Articles 25 and 28 — Transfer of personal data to third countries — Decision 2000/520/EC — Transfer of personal data to the United States — Inadequate level of protection — Validity — Complaint by an individual whose data has been transferred from the European Union to the United States — Powers of the national supervisory authorities). *Office Journal C* 398, pages 5–6.
- [Snyder, 2019] Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104:333–339.
- [Solomos et al., 2018] Solomos, K., Ilia, P., Ioannidis, S., and Kourtellis, N. (2018). Cross-device tracking: Systematic method to detect and measure cdt. *arXiv preprint arXiv:1812.11393*.
- [Sørensen and Kosta, 2019] Sørensen, J. and Kosta, S. (2019). Before and after GDPR: The changes in third party presence at public and private european websites. In *The World Wide Web Conference, WWW ’19*, page 1590–1600, New York, NY, USA. Association for Computing Machinery.
- [Straube and Fina, 2010] Straube, M. and Fina, S. (2010). *E-Commerce- und Internethrecht*, volume 5. Manz’sche Verlags- und Universitätsbuchhandlung.
- [Su et al., 2014] Su, X., Tong, H., and Ji, P. (2014). Activity recognition with smartphone sensors. *Tsinghua science and technology*, 19(3):235–249.

[Telekommunikation-Telemedien-Datenschutz-Gesetz, BGBl. I S. 1982 idF I S. 3544, 2021]

Telekommunikation-Telemedien-Datenschutz-Gesetz, BGBl. I S. 1982 idF I S. 3544

(2021). Gesetz über den Datenschutz und den Schutz der Privatsphäre in der Telekommunikation und bei Telemedien* (Telekommunikation-Telemedien-Datenschutz-Gesetz - TTDSG).

[Telekommunikationsgesetz 2021, BGBl. I Nr. 190/2021, 2021]

Telekommunikationsgesetz 2021, BGBl. I Nr. 190/2021 (2021). Bundesgesetz, mit dem ein Telekommunikationsgesetz (Telekommunikationsgesetz 2021 – TKG 2021) erlassen wird.

[the Regulation on consumer protection cooperation, 2004] the Regulation on consumer protection cooperation (2004). Regulation (EC) No 2006/2004 of the European Parliament and of the Council of 27 October 2004 on cooperation between national authorities responsible for the enforcement of consumer protection laws (the Regulation on consumer protection cooperation). *Office Journal L 364*, pages 1–11.

[Treaty of Lisbon, 2007] Treaty of Lisbon (2007). Treaty of Lisbon amending the Treaty on European Union and the Treaty establishing the European Community, signed at Lisbon, 13 December 2007. *Office Journal C 306*, page 1–271.

[Treaty on the Functioning of the European Union, 2012] Treaty on the Functioning of the European Union (2012). Consolidated version of the Treaty on the Functioning of the European Union. *Office Journal C 326*, page 47–390.

[Universal Service Directive, 2002] Universal Service Directive (2002). Directive 2002/22/EC of the European Parliament and of the Council of 7 March 2002 on universal service and users' rights relating to electronic communications networks and services (Universal Service Directive). *Office Journal L 108*, pages 51–77.

[Urban et al., 2012] Urban, J. M., Hoofnagle, C. J., and Li, S. (2012). Mobile phones and privacy. *BCLT Research Paper Series*.

[Utz et al., 2019] Utz, C., Degeling, M., Fahl, S., Schaub, F., and Holz, T. (2019). (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. Association for Computing Machinery.

[Vallina-Rodriguez et al., 2016] Vallina-Rodriguez, N., Sundaresan, S., Razaghpanah, A., Nithyanand, R., Allman, M., Kreibich, C., and Gill, P. (2016). Tracking the trackers: Towards understanding the mobile advertising and tracking ecosystem. *arXiv preprint arXiv:1609.07190*.

[Vastel et al., 2018] Vastel, A., Laperdrix, P., Rudametkin, W., and Rouvoy, R. (2018). FP-STALKER: Tracking Browser Fingerprint Evolutions. In Parno, B. and Kruegel, C., editors, *IEEE S&P 2018 - 39th IEEE Symposium on Security and Privacy*, Proceedings

of the 39th IEEE Symposium on Security and Privacy (S&P), pages 728–741, San Francisco, United States. IEEE.

[Wong et al., 2013] Wong, G., Greenhalgh, T., Westhorp, G., Buckingham, J., and Pawson, R. (2013). RAMESES publication standards: Meta-narrative reviews. *BMC medicine*, 11(1):20.

[Zang et al., 2015] Zang, J., Dummit, K., Graves, J., Lisker, P., and Sweeney, L. (2015). Who knows what about me? A survey of behind the scenes personal data sharing to third parties by mobile apps. *Technology Science*, 30.