Surveillance in the Digitalized Public Sector

A study of Third-Party Tracking on Danish Public Web Pages

Master Thesis (Kandidatspeciale) – CKOMO1045E

Date:	15/05/2020
Contract no.	15530
Line of Study	Cand.merc(kom) - 2020
Name - Student ID	Rene Meier - 101428 Teis Lebeck - 101038
Supervisor	Nanna Bonde Thylstrup
Censor	
Number of Tabs	272.639
Number of Pages	120

Danish title: ”Overvågning i den Digitaliserede Offentlige Sektor”

Abstract

Formål: De offentlige sektorer bliver verden over i højere grad digitaliseret med det formål at kunne tilbyde mere effektive, brugervenlige og nemt tilgængelige ydelser. I forlængelse heraf er den danske offentlige sektor meget fokuseret på digital forvaltning med målet om at blive en ledende aktør inden for digitaliserede offentlige ydelser. Det politiske pres for at opnå dette mål kræver dog brugen af private teknologiske kompetencer og ressourcer i form af IT-infrastrukturer (Third-party services). Det offentliges brug af disse kompetencer og ressourcer kan medføre en negativ påvirkning af borgernes privatliv, fordi private organisationers formål er at profitoptimere, hvilket også afspejles i deres brug af tredjeparts overvågning (Third-party tracking). Tredjeparts overvågning kombineret med danske borgere, som er nødsaget til at være online for at modtage disse ydelser, kan i sidste ende påvirke privatlivet negativt. Formålet med vores afhandling er derfor at undersøge niveauet af tredjeparts overvågning på offentlige danske hjemmesider ved at kortlægge det eksisterende økosystem af de tredjeparter som overvåger. Dette bidrager til en identifikation og forståelse af de underliggende faktorer i økosystemet, samt hvordan tredjeparts overvågning kan påvirke privatlivet negativt.

Teori: Den teoretiske tilgang for vores afhandling er en redegørelse for relevante koncepter inden for arenaen for tredjeparts overvågning i den danske offentlige sektor, hvor vi herved bidrager med en forståelse af de omkringliggende koncepter. Først redegør vi for udviklingen inden for digitaliseringen generelt, samt indflydelsen af big data, algoritmer og nye forretningsmodeller, dernæst for struktureringen af den danske velfærdsstat, samt hvordan tillid er en afgørende faktor i denne. Tredje del omhandler overvågning generelt, tredjeparts overvågning og forskning inden for overvågning. Fjerde og afsluttende del omhandler de mulige implikationer af tredjeparts overvågning, altså den negative påvirkning af privatlivet, den udfordrede offentlige digitale forvaltning, samt ubalancen inden for overvågning.

Metode: Vores primære metodologiske tilgang er baseret på et casestudie, da vi ønsker at undersøge det empiriske fænomen om tredjeparts overvågning på danske offentlige hjemmesider. Casestudiet af økosystemet bidrager med en dybdegående forståelse heraf, da den tillader undersøgelsen af adskillige faktorer. Casestudiet udføres ved brug af den induktive metode, da vi undersøger fænomenet tredjeparts overvågning på offentlige hjemmesider, hvortil vi kan finde sammenhænge, korrelationer og mulige forklaringer inden for dette. Vi forventer således at sige noget generelt om økosystemet for tredjeparts overvågning på offentlige danske hjemmesider. Den induktive tilgang har været mulig ved brugen af WebXray-værktøjet, som har været kilden til vores primære dataindsamling, eftersom WebXray er et *Web Crawling Tool*, som kan identificere de tredjeparter, som indsamler brugerdata og overvåger på hjemmesider. Den metodiske tilgang tager udgangspunkt i det videnskabsteoretiske perspektiv, *Sociomateriality*, hvilket tillader os at forstå teknologien med dens basale egenskaber, men også teknologien i forbindelse med forskellige aktører. Således bliver det muligt at undersøge, hvordan overvågning kan være med til at påvirke de forskellige, relevante aktører i økosystemet.

Resultater: Resultaterne af vores afhandling peger på, at der på nuværende tidspunkt forekommer omfattende tredjeparts overvågning på danske offentlige hjemmesider, hvorfor det kan konkluderes, at danske borgere er udsat for datalæk (data leak). Vores resultater peger endvidere på, at de kortlagte økosystemer er meget komplekse, og at der generelt er stor tilfældighed i brugen af tredjeparts tjenester. I færre tilfælde har vi påvist en højere grad af konsistens i brugen, hvor vi i andre tilfælde har påvist overdreven brug. Vi fandt lav til ingen indikation på eksisterende retningslinjer for brugen af tredjeparts tjenester, hvilket muligvis forklarer det omfattende niveau af overvågning økosystemet. En mulig forklaring på det eksisterende datalæk kan være et resultat af et politisk pres for at digitalisere den offentlige sektor, som har medført brugen af private kompetencer i forsøget på at opnå et tilfredsstillende digitaliseringsniveau. De danske borgeres privatliv kan som et resultat af dette påvirkes negativt, når borgerne anvender danske offentlige hjemmesider, da borgeren overvåges uden samtykke. Vi fremsætter derfor at revurdere den offentlige digitale strategi til at inkludere klare retningslinjer for brugen af tredjeparts tjenester. Denne afhandling peger altså på omfattende overvågning inden for den danske offentlige sektor, som kan medføre et pres på privatlivet. Den kan derfor fungere som springbræt for fremtidig forskning samt forskning inden for andre discipliner, såsom datalogi eller jura.

Digitalization, Politics, Public-Private partnerships, E-Government, Third-party services, Third-party services, Tracking Ecosystem, Elements, Cookies, Data Privacy, Privacy, Consent

Acknowledgements

We would like to thank our thesis supervisor, Nanna Bonde Thylstrup, for her guidance, her providing us with relevant academic materials and her role as sparring partner.

We would also like to thank Rasmus Helles for instructing us in the use of WebXray and advice for processing our empirical material.

Table of Contents

List of Tables, Figures, Ecosystems & Clusters

Tables

Figures

Ecosystems

Clusters

List of Key Definitions

Word/Concept	Definition
Web page	Online page that contains information. In the thesis, “web page” and “page” are used interchangeably
Third-party service (hereafter: TPS)	Program or service that provides third-party content for web pages. TPSs are also referred to as third-party trackers, in the context of tracking on web pages
Third-party tracking (hereafter: TPT)	Tracking of users on web pages by third-party services (third-party trackers) through HTTP requests for content
Third-party content	Content on a web page that is hosted by a third-party
Domain	A domain is an online address and identification string. Pages have page domains. Elements and cookies also request to domains, which (if operated by a third party) are called third-party domains
Element	A type of content on web pages that comes from either own servers or TPSs. Elements include, but are not limited to, images, fonts and JavaScript code
Cookie	A small piece of data, which is set in the user's browser in order to track the user's browsing activity. On web pages, cookies can be set by the web page itself or by TPSs, which are present on the web page
Consent	The user's permission to be tracked. If a user provides consent, he/she agrees to voluntarily being subjected to information collection
E-government	The process of delivering public services to citizens by using digital technologies
Governance	The process of governing and overseeing a social system through regulations, norms or laws
Ecosystem	A system of third-party trackers and web pages that contains different power structures and relationships
Ecosystem Cluster	A grouping of actors within the ecosystem, which have certain aspects in common

1. Introduction

"Rather than look for a single needle in the haystack, his approach was, 'Let's collect the whole haystack" … "Collect it all, tag it, store it. . . . And whatever it is you want, you go searching for it."

Keith Alexander – Former NSA Chief (Nakashima & Warwick, 2013)

The Digital Age is upon us, where an increasingly bigger part of our everyday lives is digitalized, tracked, and stored. The world has in the last decades undergone tremendous digital change, as we are witnessing stronger computers and therefore more data collection and processing. The Internet of Things (IoT) has on top of this led to a connectivity, making global devices capable of communicating by exchanging data at a pace more rapid than ever before (Plesner & Husted, 2019). The rise of the internet, it's infrastructure and accessibility has progressed this development even further, as it might lead to positive prospects such as freedom of expression, -assembly, and -association, which are essential values in a democratic society. Information technology has come to play an essential role in this and is more evident than ever before (Trzaskowski & Sørensen, 2019).

As a result of the digital development within business and online trade, commercialization of the internet quickly became a reality. New business models based on IT-infrastructure development started to emerge. The most recent and evident model is the "free-to-use" model, where access to the service is not monetarily based, but rather paid by "presence" on the service through being exposed to ads or through an acceptance of personal data collection. The saying "*if you are not paying, you are the product*", has become more relevant than ever before (Whitman, 2018). Organizations like Google, Apple, Facebook, Microsoft, etc., were pioneers within data collection and are superior within this business model. They instantaneously developed into what is known as data monopolies, having an extensive presence on a great part of the world wide web as we know it and are hereby capable of collecting extensive amounts of data (Uzialko, 2018).

Data as a resource through tracking has become a "currency" to be exchanged between entities or leveraged by the collector. The users of free services have become assets (Whitman, 2018). The combination of the internet infrastructure, the IoT and the extensive adoption of smart devices has led to large-scale data collection and processing of personal data. The data of today is worth more than oil, which is why some call it "*the oil of the digital era*" (Parkins, 2017). Data collection and use have several use-cases such as improving services, for prediction, for targeting, for profiling etc., and is therefore also a source of revenue. It is here interesting to reflect on the trade-off of the individual, as services are being improved and are often free to use, but data is expected as payment in the transaction for the use of these services, which might intrude privacy and result in additional implications for the individual.

Data has become essential to our society and is therefore worth more than ever before, which explains the extensive data collection that we are witnessing today. Increased surveillance of users

equals more data, which equals increased monetary value. Shoshana Zuboff has labeled this development as *surveillance capitalism*. She argues that corporations, both public and private, are collecting user data and using this for targeting purposes, which might be discriminating against the basic rights of privacy. Furthermore, these actors, who are responsible for the mass surveillance, can continue to do so, because little to none legal frameworks exist to regulate the activities of these, the world is however becoming more aware of the issue. Surveillance capitalism is based on three key dimensions, i.e. firstly, exploitation of multiple data sources through pervasive surveillance, secondly, the extraction occurs in a one-way relationship with little to no knowledge by the subject in question, and lastly complex algorithmic analytics for processing purposes allowing for complex modelling and psychographic profiling through big data analytics (Lyon, 2018; Zuboff, 2015).

The complex models and psychographic profiles made by organizations through tracking, big data and algorithmic processing can create accurate individual user profiles to be used for targeting, e.g. for marketing purposes. Dystopian consequences of these psychographic profiles could be medical insurance being based on the complex modelling of data, resulting in very expensive insurance for individuals who are genetically predisposed to certain diseases or governments or businesses using data to impact or influence populations towards an objective.

The latter has to some extent been the case. Just recently, the case of Cambridge Analytica (CA) surfaced, which was a major political scandal of tracking and stolen data without consent (Davies, 2018). CA was a data-analysis firm specialized in psychographic profiling known as targeting. CA had collected data from 50 million Facebook users without their consent through a third-party app, giving CA a monumental amount of data from American Facebook users (Davies, 2018). Several political analysts argued that CA used this data to influence the 2016 American presidential election in favour of the winning candidate Donald Trump, by targeting specific users who were undecided and influencing them through extensive propaganda posts on Facebook (Wong, 2019). Influence of entire populations as seen with the CA-case, might therefore already to some extent be a reality and is a watershed moment in the importance of consent and the fundamental right to privacy through protection of personal data (Wong, 2019). This is however a contested science, as it is contested that you can actually “provide” someone with an opinion. Privacy might be challenged with the amount of data available online, which might result in individuals being subjected to opinions that were initially not their own. Privacy is essential to a democratic society and should therefore be sustained.

Tracking and surveillance has for an extended period of time been a central topic, as Tim Libert (2015) refers to former US senators, Alan Westin and Frank Church. In 1975 they expressed concern about nationwide surveillance by the National Security Agency (NSA), claiming that this kind of surveillance could enforce tyranny towards specific groups of the people (Libert, 2015). The prediction of a society where an entity could overlook and survey the actions of its citizens, was also predicted in the book 1984 by George Orwell. The dystopian novel describes a society, in which heavy surveillance of the civilians is to ensure that the government can remain in control and uphold its totalitarian powers. Even though the book belongs in fictional literature and was first published in 1949, one could argue that it is more relevant than ever before regarding the issues of today's surveillance on the internet by both government entities and private organizations.

While tracking is a central part of today's internet use, we are however seeing a growing opposition towards tracking of data day by day, as we are witnessing campaigning for the protection of privacy (Wong, 2019). The General Data Protection Regulation (GDPR) has been implemented in the European Union (EU), which is a legal framework for regulation of data protection and privacy. The purpose of the GDPR is to protect privacy of the data subject, i.e. individuals in the EU. For companies to be able to track you in the EU, you need to consent to the tracking, giving the individual the right to opt-in or opt-out.

There are several forms of tracking, from freely giving away data hereby gaining access to services, to being tracked online from cookies. The tracking focus in this thesis is tracking performed by third-party services (hereafter: TPS) on web pages. When a user visits a web page, content on the web page, such as an image, code or font, is requested from a TPS. In order to get this content, the web page sends a hypertext transfer protocol request to the TPS, which is commonly referred to as a HTTP request. Within this request, the TPS receives information about the visitor, and then sends back the content to the web page in return. This further enforces tracking and gives the owners of the TPSs even more data for profiling and targeting, without the user knowing that this data leakage occurs.

Tracking from third-party elements are currently not covered by the GDPR, and are nearly invisible, as TPSs do not need consent to track and to collect data (Cookiebot, 2020b). In other words, TPSs can track user behaviour and movement on the Internet, with less repercussion or governance, as they are hiding in plain sight.

A single HTTP request is unlikely to result in user identification, but thousands of such requests may be correlated to a given user. Such techniques happen on the servers of corporations and are largely hidden from analysts of client-side code. (Libert, 2015:3)

Third-party tracking (hereafter TPT) in combined other forms of tracking, is problematic, since the information that leaks, allows for mapping user behaviour and actions across the Internet. A suitable analogy for TPT on web pages is a mosaic, where one could think of the information in a single HTTP request as a small piece of coloured glass. When a lot of small pieces of coloured glass are put together, they create the full mosaic. This is essentially what is happening to individual's data in TPT. The accumulated information about the same user from perhaps hundreds or thousands of HTTP requests, could potentially become user identifiable information.

As the world is becoming more digitalized, so are public sectors. Governments are becoming digitalized, to more effectively accommodate the demand from society and technological development, i.e. through improved IT-infrastructure, accessibility for users, etc. The use of digitalized technology to handle operations and deliver public services is known as e-government. Denmark is an example of a country that relies on e-government, having a highly digitalized IT-infrastructure to accommodate public efficiency. In 2018, 93 out of 100 families in Denmark lived in households with internet access. More and more Danes are using technology with internet access, and it was found that 88% of the responders in a survey from 2018, claimed that they had looked for, downloaded and submitted information via public administrative websites and online portals (Jakobsen, Jensen & Tassy, 2018). Denmark was in 2018 accredited with having the best infrastructure for e-government in the world by the UN (Digitaliseringsstyrelsen, 2018).

It is an integral part of the Danish public IT-strategy, to be a leading player in operating a digitalized public sector. The Danish government has however acknowledged that it does not possess the necessary competences and skills internally to carry out the development and implementation itself, which is why the government is collaborating closely with the Danish business community (Datatilsynet, 2016). This collaboration combined with a political pressure enforced by the government upon the public sector to become digital and to be leading, might lead to implications in the collaboration between the public and private sector. The incentive of the private sector is primarily to profit from sales, which might clash with the goal of the public sector in a welfare state, which is to serve the users most efficiently (Omobowale, Kuziw, Naylor, Daar & Singer, 2010).

The increasing dependence on a digitalized public sector, where the public is "forced" to be online to benefit from welfare services, combined with privately supplied IT-infrastructure and solutions with the primary incentive for profit, might therefore cause discrepancies in the transaction, as the level of governance might deviate (Omobowale et al., 2010).

"It is clearly problematic that public authorities are forcing us to use services where information is collected externally" (President of the IT-political association – *Jesper Lund in Boye & Bredsdorff, 2017)*

TPT has therefore become more relevant than ever before in the context of e-government, as some public web pages contain potentially sensitive personal data. When user data then leaks to TPSs, without the consent of the user, it could be harmful for the protection of privacy.

We therefore find it interesting to investigate the level of TPT on Danish public pages, and to map the current TPT ecosystem on Danish public pages, leading to our research question section.

1.1 Research Question

The primary purpose of this thesis is to present the existing nature of TPT of web pages that are owned and operated by entities in the Danish public sector, e.g. ministries, regions, etc., and to present a frame of understanding within the landscape of TPT. We are therefore by identifying the extent of TPT in a defined segment of the TPT ecosystem, able to explore the extent of data exploitation with this segment

We furthermore wish to analyze the risks and consequences of large-scale data collection in a less visible area, i.e. TPT on web pages. This allows for an understanding of how TPT creates a gap between the fundamental right to privacy, hereunder the breach of informed consent and how this might undermine democracy and personal freedom.

This leads to the following research question:

1.2 Importance, Relevance & Motivation

The following section seeks to outline the importance and relevance of this topic to the existing knowledge within the area of surveillance capitalism, and hereunder TPT.

1.2.1 Public Body with a Private Skeleton; Private-public partnerships

As aforementioned, the Danish government is categorized by having a degree of e-government, allowing citizens to use online public services for more effective processing and to search for information and communicate with public entities in an online setting. Public web pages operate on the exact same internet as private web pages, such as online stores, news portals and social media. However, given the role of the government and the public sector, public web pages arguably carry increased responsibility in the context of protecting user data. This argument is enforced by Helles, Lomborg & Lai (2019), who propose that public web pages should reduce the data collection from TPSs, because the primary purpose of these web pages is to serve the citizens in a society, rather than increasing website traffic and monetizing from this. Figure 1 illustrates of how the different sectors and spheres are overlapping.

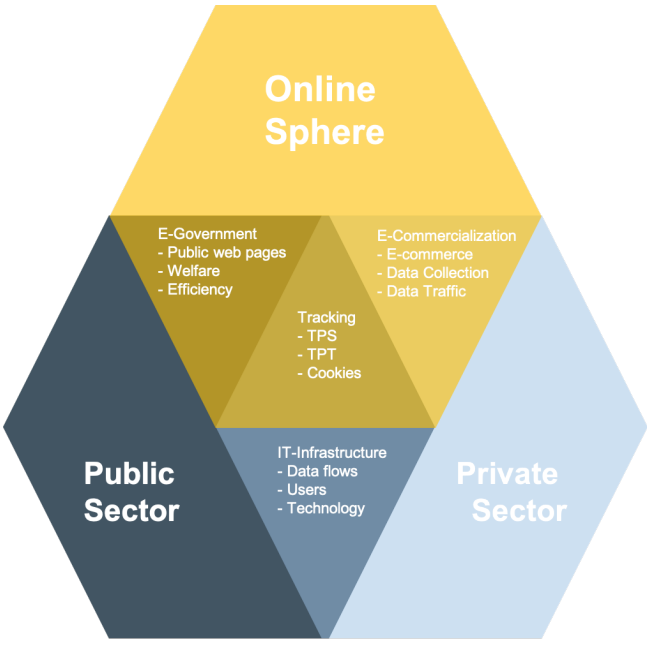


Figure 1: Sector and Sphere overlaps
Source: Own making

The requirements and demands for the Danish public sector to embrace digitalization, including e-government, is partly due to the fact that the institutions are under pressure from Danish policy and

decision makers as previously mentioned. The Agency for Digitisation (Digitaliseringsstyrelsen), an agency within the Ministry of Finance, is in charge of the Danish government's digitalization policies, including implementing digital welfare technology in the public sector (Digitaliseringsstyrelsen, 2020). In 2016, the Agency of Digitisation published its digital strategy for the period 2016 to 2020 under the name "*A Stronger and More Secure Digital Denmark*". This strategy outlines the intentions and goals of the Agency for Digitisation regarding the digitalization of the Danish public sector (Digitaliseringsstyrelsen, 2016).

The Agency for Digitalization emphasizes the wish that digital services of the Danish public sector should meet the standards of accessibility and usability of the private sector:

> *Today, we are used to seeing easy and rapid digital services from private businesses make a positive difference in our lives. For example, when we can easily transfer money to each other via our mobile phones, buy birthday presents on the internet, or check in and choose our seats in the aeroplane from home before we travel.*
>
> *It will be equally easy and quick to be in digital contact and have dealings with the public authorities. This means, for example, that the individual self-service solutions and digital solutions (such as NemID and Digital Post) have to be user-friendly, up-to-date and of high quality* (Digitaliseringsstyrelsen 2016:21).

This promise of the Agency for Digitisation is therefore a public sector matching the quality of the private sector, which could increase the pressure on the public sector to rapidly enhance its digital solutions and services. In order for the public sector to fulfill this task, the Agency for Digitisation claims that there is a need for even more public-private partnerships (Digitaliseringsstyrelsen, 2016). This is an indication that the public sector is reliant on private competences to achieve its goals. This could prove problematic or controversial given that the two parties might have conflicting interests. The public party in a public-private partnership must ensure that the outcome will provide benefits for the citizens, such as health, safety, security and efficiency. However, the private party involved is likely to be dominated by monetary interests (Omobowale et al., 2010).

A current example within the Danish public sector, is the relationship with Singularity University, which is an American think-tank and consultancy firm in the technology sector. The founder and CEO of Singularity University, Ray Kurzweil, who is also the director of engineering at Google, is a significant proponent of technological singularity; A hypothetical point in the future, where technology becomes so strong that its advancements will be irreversible and transform human life (Cadwalladr, 2014). In 2016, Singularity University was invited by then Prime Minister of Denmark, Lars Løkke

Rasmussen, to speak about disruption at an inspirational meeting. This inspirational meeting was a part of the initial processes of establishing the Disruption Council; A council formed by the Danish government, consisting of members of parliament, leaders of labour unions and CEOs of large, private companies (Statsministeriet, 2017).

Seeking inspiration and guidance from a private organization, like Singularity University, could potentially impact future decisions and policies within this area, due to the significant difference in ideological beliefs. This concern was exemplified in 2019, when the chairman of The Business Committee (Erhvervsudvalget) in the Danish Parliament, Lars Christian Lilleholt, directed a question towards the Ministry for Industry, Business and Financial Affairs, as to whether the ministry had used courses and workshops from Singularity University as a source of inspiration for the formulation of digitalized legislation (Finansministeriet, 2019). The Minister of Finance, on behalf of the Agency for Digitisation, denied this, even though Singularity University was invited to Denmark by the government.

The authorities may do so in a good sense because the purpose is to improve their services. But they pay with the privacy of the citizens (Associate professor at IT University, Thomas Hildebrandt, 2017)

It is clear that the public sector wishes to be innovative and effective in the context of digitalization. However, for this to happen, private actors must be included in the process. We seek to understand if the fast-paced digitalization of the public sector, with private actors playing a central role, could be compromising for the rights to privacy on the internet. Given the size and scope of the digitalization of the Danish public sector, it is a reasonable interest to ensure compliance with current and future legislation.

In 2017, the Danish Business Authority (Erhvervsstyrelsen) and the Agency for Digitisation published an official guideline explaining the policies for web pages of public authorities. In the document, it is stated that the Danish Business Authority in 2016 analyzed the respective web pages of 60 Danish municipalities and 12 public self-services. Based on the results, which are not specified, the Danish Business Authority deemed that there was a need to create a guideline for the use of cookies (Erhvervsstyrelsen, 2017). The guideline specifies that third-party cookies from social media, such as Facebook and Twitter, should only be activated when the user clicks on that element on the web page. Furthermore, the guideline specifies that users should be able to access self-services and basic information, without being exposed to cookies from TPSs and that it is the responsibility of the

individual public entity to ensure proper governance of the third-party cookies on their web page (Erhvervsstyrelsen, 2017). While the document is centered around tracking from third-party cookies, there is no mention of the fact that these plugins still receive information about the user's IP-address and so forth, even without the use of cookies. This could be an indication that more tracking is happening than what is known by the Danish Business Authority. This further adds to the relevance of our thesis, as we have already seen some coverage on TPT on Danish public web pages, but we believe that there is more knowledge to be added to this topic.

In a more recent case from February 2020, the Danish Data Protection Agency criticized the Danish Meteorological Institute (DMI), part of the Ministry of Climate, Energy and Utilities, in a judicial decision, for not processing user data in compliance with the GDPR. The criticism was rooted in the fact that DMI did not obtain the users' consent to share their information, i.e. forwarded IP-addresses and activities on the web page to Google through banner advertisements on DMI's web page (Datatilsynet, 2020). The agency found that the information, which consisted of the IP-address, name of website, type of browser and time of visit, could be characterized as personal information, as it could be used to target individual users with personalized advertisements. The agency claimed that this information could be retrieved from cookies. DMI changed the opt-out function and made it easier for the user to reject cookies when entering the website. While rejecting cookies could lead to more privacy, this type of information could also be retrieved through tracking from third-party elements on web pages, which does not require the user's consent. This is why TPT is problematic, and also leaves the question on whether the current TPT should also be considered as noncompliant with the GDPR.

Public-private partnerships as stated above might therefore cause incompatibility between the digital agenda of the protection of the individual, as private parties and public parties have a conflict of interest. We therefore find the subject of TPT on Danish public web pages both important and relevant to study.

1.3 Scope & Delimitations

The following section is an account of our focus throughout the thesis. We are here accounting for our delimitation choices, as well as offering a description of the different approaches and decisions that have been taken.

1.3.1 Specific Arena of Tracking

This thesis will be undertaking a national focus, hereunder a sole focus on the Danish Public web pages. This focus has been chosen, as the Danish government wishes to be seen as a leader in delivering public services through digital channels, also called e-government. In Denmark, the level of welfare is seen as high, compared to the general benchmark of welfare, as the majority of public services are free due to the entirety of the population paying an increased level of taxes (OECD, 2019).

This allows for higher quality of security, healthcare, education and general public services. The increasingly digitalized public sector allows for more efficiency within public services, which as a result should further improve the welfare of the state. We are therefore going to be focusing on the digital public arena in Denmark, as tracking within this arena could potentially have consequences for the privacy of the individual. This leads to the next section, which is the type of tracking that we focus on.

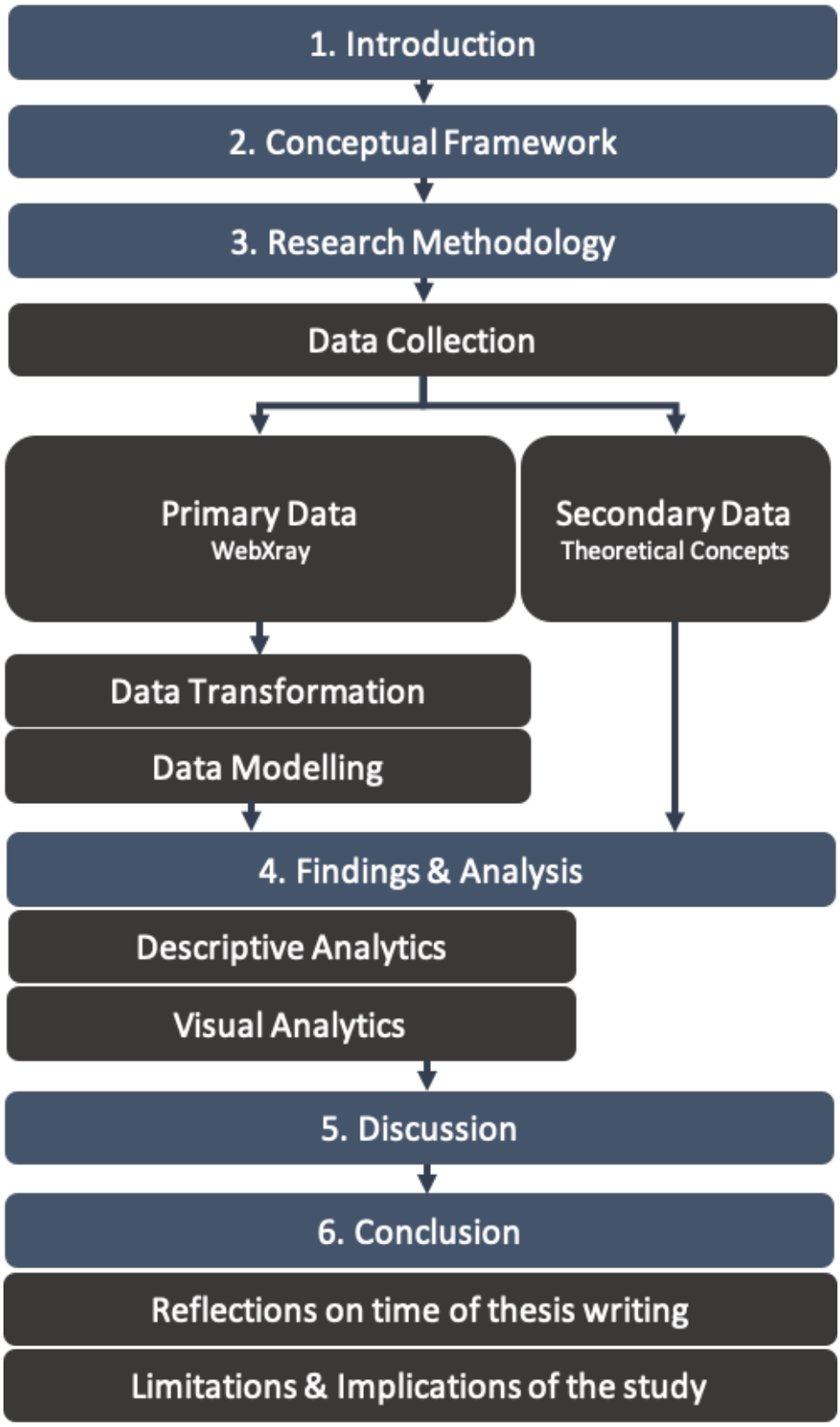
1.3.2 Specific Type of Tracking

Online tracking is the practice of collecting data about an individual's online behavior. Several different methods of tracking exist, e.g. freely given data to use a specific service, cookies on web pages, analytics in general, etc. (Cookiebot, 2020a). All these trackers are governed by the GDPR, meaning that consent is necessary for these trackers to collect data. We have chosen to focus on third-party trackers, as these are not governed by the GDPR. Third-party trackers are services that track users through requests from web pages for elements, such as images, fonts and code.

This form of tracking does however not rely on consent and is therefore capable of collecting data without permission of the individual. We find TPT interesting in relation to the increase in the digitalization of the public sector in Denmark and hereunder the e-government, as the public is “forced” to use these online services and are therefore subjected to tracking beyond their consent. Some public web pages might contain sensitive information, such as medical information, which could furthermore harm the fundamental right to privacy of the individual.

We have therefore limited our thesis scope to TPT on the web pages of the Danish public sector. The scope of our thesis is therefore to map a fraction of the entire ecosystem of third-party trackers, hereby being able to say something general about the ecosystem of third-party trackers on Danish public web pages.

1.4 Thesis Structure



Part 1 is the introduction to the subject of TPT on Danish public web pages and a clarification of why this an important and relevant field of study. After having introduced a relevant field of study.

Part 2 is the conceptual framework we are using throughout the thesis and is an elaboration of our introduction. We are here going to explain and discuss the relevance of different concepts being important to the understanding of the study, as well as explain the application of the different concepts. Part 2 will offer an understanding of different subjects such as privacy, consent, tracking, TPT, digitalization, the digitalized state, data-based business models, challenged governance, etc.

Part 3 is the methodology of our thesis. We are here accounting for our research methods, such as data collection, data transformation, data analysis, data visualization, data understanding, limitations of the data, etc. We are therefore in part 3 going to explain in detail how we have conducted our study, to ensure validity and reliability of the thesis.

Part 4 is where we account for our results, as this section is the "findings & analysis"-section. We are here going to present our raw findings, and afterwards analyze these findings, to identify possible patterns, relations, and other interesting points. The combination of part 3 and 4 is therefore going to cover the empirical study of the thesis.

This leads to **Part 5**, which is a discussion of our findings. The purpose of our discussion is not to include new findings distant to our current analysis, but to account and connect all our findings to construct a bigger picture of TPT on Danish public web pages. We are here going to further elaborate the consequences and implications of our findings.

Part 6 is the conclusion of our thesis and is the answer to our stated research question. The answer will be based on the analysis in part 4 condensed to the most important findings in part 5, the discussion. Part 6 also contains 6.1, which is a broader perspective of our thesis in the midst of the Corona-crisis. We are here discussing technology in this context. Lastly, part 6.2 is about the limitations and possible future research of our study.

2. Conceptual Framework

“Surveillance capitalism claims human experience as raw material for translation into behavioral data. That data is partially used to improve the digital products or services; but most importantly it is declared "proprietary behavioral surplus’" fed into "machine intelligence" manufacturing processes producing ‘predictions products’". These "behavioral prediction products" are sold in a new type of market: the "behavioral futures market"

Shoshana Zuboff (Zuboff, 2019:8)

This section is a description of theoretical concepts within the arena of tracking to offer a better understanding of the field and to describe the relationship between the different concepts. It is furthermore a description and rationalization of the application of the separate theories, i.e. an explanation of why we have chosen to include these theories under the specific sections and how we are applying them. Some of the theories and concepts explained will not be practically applied after the conceptual framework section, as they are serving as an understanding of the concepts of the thesis.

2.1 The Age of Digitalization

We are currently living in the age of digitalization, as an extensive part of our lives are being digitalized, allowing for tracking and measuring. To provide an understanding for this development, we are going to be drawing on the theories by Plesner and Husted (2019) hereby discussing the differences between *Digitalization, Datafication* and *Digital Transformations* and the relevance of these concepts to our study. The current and increasing digitalization has led to increased tracking of several phenomena from stock market prices, weather forecast, etc., but also tracking of private individuals. Some organizations are profiting from this increased digitalization, which is why tracking has become relevant in business aspects. This has come to be known as *Surveillance Capitalism.*

The increasing data dependency caused by central players such as Google, Facebook and Amazon, has led to an exponential increase in tracking of private individuals. We are drawing on theories by Zuboff (2019) and Lyon (2018) to gain an understanding of why the concept of surveillance capitalism has become evident and the underlying causes of tracking within surveillance capitalism. The digitalization and surveillance capitalism have been made possible by *big data*-analysis and *algorithmic transformation*. We are therefore drawing on theories by Gillespie (2014), Boyd & Crawford (2012), Flyverbom & Madsen (2015), and Mayer-Schönberger and Cukier (2013) as these authors offer an understanding of the concepts of big data and algorithms. They offer an understanding of how big data and algorithms have come to play a big role in society and what the challenges and implications are, as well as how tracking has been made possible through these concepts.

This section is therefore an account and explanation of how surveillance capitalism and hereunder tracking has become evident in our current society and how it is made possible by the digitalization, big data and algorithmic transformation of data.

2.1.1 Surveillance Capitalism

Our current society is unavoidably becoming increasingly surveillant due to an increasing data-dependency. Surveillance is seen in nearly every aspect of society, e.g. social interaction, government interaction, location interaction, etc. Surveillance is becoming part of our everyday life (Lyon, 2018). The development within surveillance is by no means an innocent development, as the development is led by global organizations with a monopolistic status in data collection. This new form of surveillance has come to be named "surveillance capitalism", as surveillance has been capitalized (Zuboff, 2015).

Wielding the power of surveillance data is to possess immense power, control and profitability according to Zuboff (2019). The development from governmental surveillance of protecting its citizens to surveillance capitalism where organizations are surveying for profitability is by no means an odd development. This is why we are witnessing every part of society from healthcare and educational establishments to politics and small business owners wanting to be part of the big data development, as the internet giants have shown the potential within surveillance data (Lyon, 2018; Zuboff, 2015).

A prime example of a surveillance capitalism organization is Facebook. You are not paying to use the services of Facebook, but how do they make money then? The business model of Facebook is about connecting users with other users, to create specific networks of people. These networks are then analyzed through the enormous amounts of data left behind when using different services, to create psychographic group profiles, but also very specific individual profiles. The success of this profiling is based on different kinds of tracking. Facebook does as well connect users with unseen others, i.e. data brokers, data vendors, advertisers, developers, political campaigners, etc., who are paying well to access this kind of data due to the monopolistic nature and complexity of the collection. The business model of Facebook is essentially interactions with the platform, which is why people are encouraged to spend time on the platform, allowing for increased data collection and in the end prediction of lifestyles (Lyon, 2018).

When asked about privacy concerns for the users, former data manager for Facebook, Sara Parakilas, said:

> "*Facebook prioritizes the growth of users, the growth of the data they can collect and their ability to monetize that through advertising…those are the metrics that the stock market cares about*" (Stahl, 2018).

Shoshana Zuboff is a pioneer within the concept of surveillance capitalism and already in 1988 published "*In the Age of the Smart Machine*" where she argued how technology allows for increased transparency and knowledge, but also increased surveillance of the workers. One of her most recent additions "*The Age of Surveillance Capitalism*", is about the emergence of commercialized surveillance. The book argues how internet organizations through surveillance business models have obtained massive valuations, e.g. Google $600B, Apple $750B, Microsoft $521B, Facebook $420B, etc. The "secret" to profitability as argued by Zuboff is what she calls "unilateral surveillance and behavior modification" (Zuboff, 2019), which is about selling real-time access to people's everyday life through targeting and changing behaviors. We are becoming increasingly dependent on the digital infrastructure in our everyday lives, allowing for the straight extraction of consumer data, without any transaction with the users. Personal data has become an asset to be traded between organizations (Zuboff, 2015).

Surveillance capitalism differs from traditional capitalism, as traditional capitalism is based on supply and demand, whereas Surveillance Capitalism is based on endless accumulation, meaning more is better. The key dimensions of surveillance capitalism are firstly, exploitation of multiple data sources through pervasive surveillance, secondly, the extraction occurs in a one-way relationship with no to little knowledge by the subject in question, and lastly complex algorithmic analytics for processing purposes allowing for complex modelling and psychographic profiling through big data analytics (Lyon, 2018; Zuboff, 2015).

We are undoubtedly witnessing a successful capitalist development with surveillance capitalism, as the highest valued organizations in the world are the ones leveraging the power of data through tracking. Behavioral data has become an asset, as our social sphere is controlled by organizations. The question raised by experts is however, if we wish to be subjected to few organizations possessing this kind of power? (Lyon, 2018, Zuboff, 2019).

These data monopolies are leveraging the data without users knowing the true purpose. This is where the concept of *usage and abusage* becomes relevant. Just recently, the case of Cambridge Analytica (CA) surfaced, which was a major political scandal of tracking and stolen data without consent (Davies, 2018), as mentioned in the introduction.

The development of surveillance capitalism has been made possible by the rapid digitalization we are witnessing right now, allowing for big data analysis augmented by algorithms, which will be explained in the next two sections.

2.1.2 Digitalization, Datafication and Digital Transformation

The world is, as previously introduced, undergoing significant digital changes. We are every day witnessing how the world is becoming increasingly digitalized. Within the industrial sector, organizations are offered greater flexibility, reactivity and product individualization. Complexity has on the other hand increased as well due to this rapid change and demand of adaptation. These digital changes are often referred to as *digital transformations*. Digital transformation is however a vaguely describing word, as it consists of several factors and is dependent on the context of use. Digital transformation consists of two separate parts *Digitalization* and *Datafication.*

Datafication is essentially transforming analog information into data to be used in a digital context, i.e. allowing computers to store the information that was previously stored physically, e.g. paper. Datafication is defined as “the process of changing from analog to digital form” by the research and advisory firm Gartner (Bloomberg, 2018). The earliest form of datafication is converting handwritten text into digital form. This process has however developed exponentially in the last few decades and potentially everything can be datafied from text to the movement of individuals. It is however important to emphasize that information is datafied and not the process, as this is where digitalization becomes relevant (Bloomberg, 2018).

Digitalization is on the other hand a more complex entity and is defined as “the way in which many domains of social life are restructured around digital communication and media infrastructures” by J. Scott Brennen and Daniel Kreiss from the University of North Carolina or as “the use of digital technologies to change a business model and provide new revenue and value-producing opportunities” by Gartner (Bloomberg, 2018). The Gartner definition is more business based rather than academically founded. The two quotes vary in clearly defining digitalization but do have common ground in the implementation of digitalization technologies. Digitalization is as well closely related to automation, as the purpose of implementing information technologies usually is to improve efficiency and increase data transparency (Bloomberg, 2018).

The datafication of information and digitalization of processes has therefore resulted in the digital transformations we are witnessing today. The end product is an increasingly automated world, with

more data to be leveraged for increased efficiency. Increased digitalization does however rely on increasingly complex IT-infrastructures. IT-infrastructures are defined “large systems of hardware, software, and networks that are installed in organizations to support their activities (Plesner & Husted, 2019:16).

The complexity has increased within data flows, i.e. the complexity of how data is flowing, how data is being transferred, how data is transformed, etc. Modern organizations like Google, Facebook and Amazon have extraordinarily complex and complicated IT-infrastructures and flows of data. These organizations are using a combination of on-premise computing and storage and off-premise cloud-based resources (Myers, 2018). These complex systems combined with the IoT have led to an even more complex system of interconnected systems, to allow for the unification of data in one place, which from the illustration below is the middle field called “Data Consolidation / Application”.

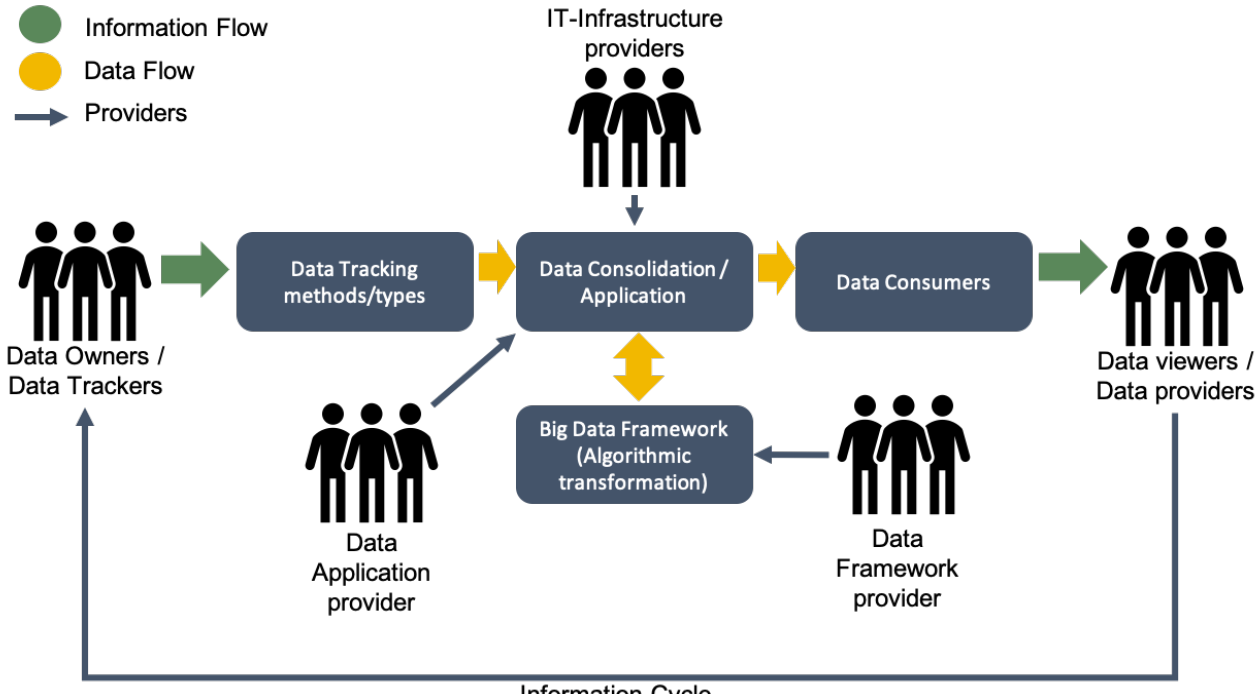


Figure 2: Information cycle
Source: Own making with inspiration from (Wong, 2019)

Figure 2 illustrates a broad overview of how data flows and the roles of different actors in the information cycle. The data owners/trackers formulate a request and purpose of tracking, which is illustrated as information flow. The purpose is then fulfilled by using different tracking methods, which feeds into the data consolidation center. The data consolidation center then requests the *big data framework* for transformation of the data through an algorithmic process, allowing for usability of the data. When the data is handled by these complex processes it is then made available to

users, who then uses the service, thereby giving away more data, and the process starts over in an endless cycle.

This is a very simplified overview of the process, as the underlying IT-infrastructure and data processing tools are very complex and depend on a wide range of skills and competences, meaning it is a very costly process. Few organizations master the full complexity of these infrastructures, but those who do, possess a great advantage. An advantage great enough for academics and policy makers to debate whether the control over data in such scale confers unfair competitive advantage, as stated in a report by the European Policy department for economic and scientific policy (European Parliament, 2017). The argument from the report is that the complexity within IT-infrastructures that allows for large amounts of data, equals abuse of a monopolistic position within the market, and it is therefore argued that those who control large amounts of data should be required to share it (European Parliament, 2017). The measures discussed at the moment, are blocking mergers, which further enhances the position of the data monopolists, prosecution of organizations keeping data from competitors and banning platforms who collect an extraneous amount of data beyond customer's direct requirements (European Parliament, 2017)

2.1.3 Big Data and Algorithmic Transformation

To achieve a better understanding of the algorithmic processes mentioned above in the data consolidation / application process, it is here relevant to draw on the definition of Gillespie (2014), who defines algorithms as specified calculations capable of transforming input data with the help of encoded procedures, i.e. allowing for organizing disorganized data (Gillespie, 2014:1). Algorithms used in data transformation range from very simple to very complex. Complex algorithms are based on specified calculations and encoded procedures (Gillespie, 2014). Algorithms are however not some "magic box" turning raw input into useful insights. Algorithms need to be connected to a machine / database providing the information for the algorithm to transform and are reliant on human coding (Gillespie, 2014).

Algorithms as previously stated depend on input that allows these specified sorting processes to transform raw input into useful insights. This is where the concept of big data becomes relevant. Big data has become a buzzword in society and is defined as "*speedy ways of compiling, combining, and mining multiple types of data, rather than looking for singular evidence*" (Flyverbom & Madsen, 2015:884). Big data analysis allows for analysis of very large data sets gaining insights that were previously impossible to achieve. These datasets offer an aura of truth, objectivity and accuracy, as

the analysis allows for the “objective” identification of correlation between multiple sources of data (Boyd & Crawford, 2012). This kind of analysis has therefore come to be known as big data analysis, as it is the analysis of vast amounts of real-time sources (Flyverbom & Madsen, 2015). According to Mayer-Schönberger & Cukier (2013), the era of big data is still in its initial phase, as we are first starting to see the capabilities of the technology and therefore the benefits and consequences as well.

Big Data according to Flyverbom & Madsen (2015) is seen as “*a homogenous phenomenon that will disrupt a range of established ways of working, living and thinking*” (Flyverbom & Madsen 2015:126). Big data allows for extreme precision in targeting, due to the vast amount of data, as the analysis is a compilation of many different sources, as it is the case with tracking and targeting. Organizations who leverage big data analysis by having access to different sources are capable of producing to a high degree accurate psychographic profiles of individuals.

The processing of vast amounts of data is usually augmented by algorithms based on unique mathematical methods fit to a certain context, meaning that the processing of data is augmented by human coding. This is the only way for the data to be turned into actionable insights (Mayer-Schönberger & Cukier, 2013). This raises the question of objectivity, as the data cannot be categorized as neutral, neither can it be categorized as “found”, as the algorithm has been programmed by humans to fit into an organizational context, which equals the “request” mentioned in the above figure (Flyverbom & Madsen, 2015). Data is framed, but is also framing, meaning that the useful insights have been framed for a specific context but also further frames the context (Flyverbom & Madsen, 2015). Big data is messy and vast due to the reliance of several real-time sources and it therefore relies on correlation rather than causation, as it looks for patterns in data, and not reasons for patterns. Messiness combined with the identification of patterns might lead to untrue correlations, enabling the practice of apophenia, i.e. seeing patterns, where none exist. It is however argued that the amount of data balances the pattern errors (Boyd & Crawford, 2012; Mayer-Schönberger & Cukier, 2013).

Big data is viewed as “*a source of new economic value and innovation*” (Mayer-Schönberger & Cukier, 2013:12), which is consistent with the concept of surveillance capitalism, as these organizations rely on data for the purpose of improving their business model and to profit from the data.

The combination of big data as analysis tool and algorithms as processing tool, therefore, allows for specific psychographic profiles. The analysis does however rely on human coding, raising the question of objectivity and should according to Flyverbom & Madsen (2015) not be perceived as truths due to the correlation vs. causality balance of the analysis.

2.2 The Digitalized State

As we seek to understand the current extent of TPT on Danish public web pages, it is important to offer an understanding of the role of the public sector in Denmark. To understand the roles of the branches within the Danish government and public authorities in terms of delivering services to its citizens, we rely on the work of Greve (2018), Vrangbæk (2009) and Greve, Lægreid & Rykkja (2016). This section covers the concept of the e-government, as to why it is necessary to understand this concept and the enabling factors. In terms of defining the scope of e-government, we work with the definition presented by Rey-Moreno, Felício, Medina-Molina & Rufín (2018) where e-government is defined as delivering public services through digital channels. For describing the enablers of e-government, we draw on theories by Forrer, Key, Newcomer & Boyer (2010) hereby offering an understanding of the need for public-private partnerships, and thus the inclusion of private actors, in order to realize certain projects in the public sector.

We draw on theories by Jensen & Svendsen (2009) to explain the mutual trust between the government and citizens in Denmark on a general level and from Lauritsen (2011) to explain how the Danish government's ability to collect and handle personal data about the citizens is enabled by the mutual level of trust.

2.2.1 Structure of the Danish Welfare State

The central political size of the welfare state in Denmark is a key driver for the digitalization and datafication of the public sector. The type of welfare system in Denmark is a *Nordic welfare system* (Vrangbæk, 2009), which is also known as the *Nordic model*. Welfare states of the Nordic model are characterized as universal welfare states with active labour market policies and overall high level of equality (Greve, 2018). However, since the welfare system is universal and thus needs to serve a large group of people, it needs to have an adequate public sector that can manage and deliver these services. The responsibility of delivering public services in Denmark is divided into different authoritative levels, i.e. state-, regional- and municipal levels. This division is based on the work of Vrangbæk (2009) and the Agency for Digitisation's outline of the Danish public sector (Digitaliseringsstyrelsen, 2019).

On the state level, there are mainly the ministries and state organizations. Within the responsibility area of the ministries lie both ministerial departments and ministerial agencies, such as the Agency for Digitisation, which lies under the Ministry for Finance. The state organizations are also located within the responsibility areas of the ministries, such as law enforcement lies under the Ministry of Justice. In Denmark, compared to the other Nordic states, the ministries hold a very strong position in terms of responsibility (Greve, Lægreid & Rykkja, 2016).

Whereas the state level operations are nationwide, the regional level covers the areas that are delegated to the five different administrative regions of Denmark. The main responsibility areas for the regions are health, with hospitals, emergency rooms and psychiatric institutions all being administered and operated by the individual regions.

The last and lowest level in terms of hierarchy, is the municipal level. There are currently 98 different municipalities across Denmark, which differ largely in size, both in terms of population and area. The municipalities also control certain entities within their own area, such as primary schools and high schools.

2.2.2 Trust is the Key to the Welfare State

One key element allowing the Danish public sector to retain its large size, is the high level of trust that exists in society. In Scandinavia, the level of social trust, which is the idea that the majority of the people can be trusted, is relatively higher compared to other European welfare states. This social trust allows citizens to willingly contribute to the common good and support the large public sector (Jensen & Svendsen, 2009). Though social trust mainly describes the relationship between people (Jensen & Svendsen, 2009), a strong mutual trust between government and people also exists in Denmark. This level of trust is important, as it is not necessarily the same elsewhere in the world.

Literature surrounding the topic of government surveillance and data have come from other countries and cultures, where the level of trust is not the same as in Denmark. As surveillance relates to Denmark, Lauritsen (2011) argues that surveillance is necessary for the Danish society, because it is a central part of its infrastructure. However, though the notion is that surveillance is central to society, there are different opinions on how to approach it (Lauritsen, 2011). An example of Danish government surveillance is the CPR-number, which is the Danish civil registration number that was

implemented in 1968. The initial incentives for creating the CPR-number was to get a better overview of the citizens, because it would be necessary to create a more efficient and fair taxation system.

One key issue for the government, when implementing the CPR-number, was to define how much information about the individual citizen, which the number should contain. Lauritsen (2011) refers to this as an *information theoretical dilemma,* where information about the individual would give the government and public authorities a better overview, whilst also potentially leading to identification of the individual. Ultimately, the CPR-number ended up containing information about date of birth and gender. From this dilemma, parallels can be drawn to the issue of TPT on public web pages. Here, there is also the issue of identifying the user from the information in the HTTP requests to TPSs (Libert, 2014). However, though the Danes are aware that the CPR-number contains information that could be abused in some way or form, they still do not consider the CPR-number as something problematic:

> *"One might know that the CPR-number can contain certain information, which theoretically could be abused. However, we principally do not have anything against the fact that the state has access to information about us. We are confident that the information is treated carefully and in accordance with all protocols, and we perhaps believe that giving up personal information is a condition for maintaining an efficient welfare state with healthcare, social services and much more"* (Lauritsen, 2011:13).

Whereas this point from Lauritsen (2011) describes the willingness to give up personal information to the state, it does not describe a situation in which an individual gives up personal information to a private entity, such as a TPS. Some elements and cookies from TPSs could contribute to maintaining an efficient welfare state, such as making information on public web pages more easily accessible. However, there might also be some third-party content that is not central to the purpose of these web pages, which raises the question whether the people would accept the data leak to these TPSs, if the third-party content is not strictly necessary.

2.2.3 E-government & Public-Private Partnerships

The birth of the CPR-number was due to the Danish government's wish to get a better overview of its citizens, in order to create a more efficient taxation system. Though the trust between people and government played a big role in making it possible, the development in information technology was also a key contributor (Lauritsen, 2011). As information technology has developed further and digitalization has become a political agenda, states have found new ways to deliver information and services to its citizens on digital platforms. This is referred to as e-government and can be defined

as the practice of delivering information and services to citizens through digital channels (Rey-Moreno et al., 2018).

Literature surrounding e-government suggests that the implementation of it started around the beginning of the 21st century. One such example is from the United Kingdom, where the then sitting government in 1999 announced that they intended to deliver all of its public services electronically by 2008 (Computer Fraud & Security, 1999). At the time, more Brits thought that the potential of fraud and computer literacy would be the biggest barriers for e-government, while only a very small group thought that the credibility of the private third parties would be problematic (Computer Fraud & Security, 1999).

Information technology has developed further, and an increasing number of Danes are using digital technologies in their everyday lives (Jakobsen et al., 2018). Furthermore, there has been an agenda in Denmark to modernize e-government to an extent where it is equivalent to the level of digital solutions seen in the private sector (Digitaliseringsstyrelsen, 2016). However, for any state to make this happen, the state must rely on the skills of the private sector and third parties.

> *"In a globalizing world that is more integrated, complex, and volatile, governments simply may not possess the prerequisite knowledge, capacity, or managerial skills. When this is the case, governments need to engage partners that have the necessary expertise, know-how, and managerial adeptness needed to carry out government responsibilities"* (Forrer et al., 2010:477).

Given the complexity of digitalization, states and governments have sought to include private actors to aid them in building digital infrastructures. Establishing IT-infrastructures are essential for some parts of the public sector, where digital transformations are absolutely needed. This can be seen in areas such as horticulture, where an example from the Netherlands has shown that new digital structures have been able allow the government to better align operations and execute strategies (Verdouw, Bondt, Schmeitz & Zwinkels, 2014). In the context of being reliant on private actors, it is both in terms of creating the infrastructure itself, as well as the training of employees in the public sector to use the new technologies and solutions (Ogonek & Hofmann, 2018). The importance of preparing employees in the public sector for digitalized solutions is also a focus by the Organizations for Public Employees (OAO) in Denmark, claiming that training of employees must be a central task in the process of digitalizing the public sector (OAO, 2020). Though the inclusion of private actors seems highly necessary for Danish public sector to reach a level of digitalization matching that of the private sector Digitaliseringsstyrelsen (2016), the dilemma of giving away control persists. This

essentially becomes a question of accountability, on whether the public party should still be held accountable for what might happen in the processes undertaken by the private party. Though the public party is not performing the same tasks as the private party, the public party, such as the Danish state in this case, is ultimately responsible for the outsourcing (Forrer et al., 2010).

2.3 Tracking

In order to gain a better understanding of TPT on web pages, we have included literature providing a historical outlook on the studies within the concept of tracking. We draw on theories by Agre (1994) to offer an understanding of the difference between "surveillance" and "capture" in a timely context and Krishnamurthy & Wills (2006) to offer an understanding of the beginning of TPT on the internet, as they offer some of the earliest studies on TPT. However, as the research methods of this rather complex concept have been made more available for researchers outside of the field of computer science, where there previously was a gap between academic disciplines, we draw on the works of Bucy & Zelenkauskaite (2016), who argue that the complexity of concepts related to big data have forced this gap, and Libert (2015), who claim that new tools for researchers have sought to bridge this gap.

We share the view of Libert (2015) that studying TPT on the internet should include a focus on the social implications of this type of tracking. Discrimination is a social implication from TPT, which is why we include the works of Sweeney (2013) and Libert (2014). Both authors include cases where collection of information about users on the internet led to discrimination towards certain users based on demographic characteristics. To better understand the method for studying TPT that will be used in this thesis, we draw on Libert (2015) and Helles et al. (2019), as both have used this specific method.

2.3.1 What is Tracking?

Tracking on the internet can be defined as a practice that "*gathers user data to perform online advertisement, content personalization or user authentication*" (Sanchez-Rola, Ugarte-Pedrero, Santos & Bringas, 2017:18), and this can occur in many different ways. Even when we narrow down the tracking to digitalized and datafied contexts, there are still various types. Tracking based on location occurs when a user is carrying a device, such as a smartphone with enabled location based services, i.e. services requiring access to the geographical location of the user (Lehrer, Constantiou & Hess, 2011) This is the case with mobile applications such as Google Maps, Uber and Endomondo. Facial recognition has also been used for tracking purposes, such as in the social credit

system in China, albeit this type of tracking is more analog. The Chinese government has installed more than 200 million cameras with facial recognition software, which can record and recognize citizens that commit minor offences (Marr, 2019). If a citizen is recognized while committing the offence, he or she will automatically be registered.

The tracking of users is not restricted to one platform, as some organizations practice tracking methods on various platforms in order to gain information about the same and single user. Law enforcement agencies are examples, as they have been using video material from CCTV and GPS locations from Google Maps, in order to identify suspects (Shih, Chen, Cheng & Kao, 2019). Though CCTV is a type of surveillance that is not datafied, it is augmented by the data from Google Maps. The social credit system in China is also not only gathering data points from facial recognition from camera surveillance, but also from social media activities and online shopping (Marr, 2019). Another type of tracking, which is the type that our thesis is centered around, is tracking performed by TPSs on web pages, i.e. TPT. TPT will be covered in the following section.

The Importance of Considering and Studying Tracking

The idea of studying human behaviour with the purpose of optimizing advertisements and market offerings is not new. For decades, marketers have studied existing and potential customers to provide themselves with a stronger knowledge of what to offer and how to do it. What sets apart tracking based on user data from the internet, is that it has reached a scale where one could argue that companies are no longer just observing users, but rather surveilling them. This is what Zuboff (2019) calls surveillance capitalism, as the accumulation is endless.

One key argument in the critique of tracking on the internet, is the idea that identifying users through tracking, results in discrimination of the same users. Latanya Sweeney (2013) found, in study conducted in the United States, that a user was more likely to be shown advertisements related to criminal records databases when searching for popular African American names, compared to popular Caucasian American names. In another case, an American data broker sold a list of users characterized as “rape sufferers”, “domestic abuse victims” and “HIV/AIDS patients” (Libert, 2014). Even though that these examples are from the United States, there is no indication that such discrimination could not or is not already happening in Denmark. We work from the notion that if a TPS receives any information about the user, then this information has potential to be used to target the user.

2.3.2 What is Third-Party Tracking

Tim Libert (2015) describes how a TPS can retrieve information about the user on a given web page. There are mainly two ways that the third-party can retrieve information about the user; third-party elements and cookies. Aforementioned, third-party elements, such as images, fonts or code, are loaded into a web page, when a user enters it. In return, the TPS that provides the content will receive information about the visiting user.

The information is included in the request for content that the web page sends to the TPS. This request, which is the HTTP-request, contains information about the user, which is the IP-address, the type web browser and computer, date and time for the request (Libert, 2015). Cookies are another type of tracking mechanism, which is set in the browser of the user in order to track the user's behaviour through the browsing session and to recognize if the user is a recurring visitor (Libert, 2015). Cookies can be necessary for certain third-party content to function, such as Google Analytics.

While none of these types of TPT can see the identity of the user, gathering data from multiple web page visits throughout the internet could likely reveal who the user is, which creates the danger of both user identification and discrimination (Libert, 2014). This concern is also shared in Danish public sector by the Danish Data Protection Agency (Datatilsynet), as to why it is possible that this type of tracking is noncompliant with the personal data security legislation. According to the Danish Data Protection Agency, the severity of a breach is based on whether the leaked information could provide personal information about the user and thus potentially create harm:

> *"On one side, the disclosure of a small amount of very sensitive personal information could do great damage, while on the other side, it could similarly have severe consequences if many pieces of information put together could harm the affected individual"* (Datatilsynet, 2018:8).

This is relatively similar to the dangers Libert (2014) refers to and could further support the argument that tracking by TPSs on web pages does not fully comply with the current legislation on this area. The argument that many data points about an individual user can lead to identification, is enforced by the ecosystem of TPSs on the internet. Though there are many different TPSs, where some of them only operate on very specific web pages (Helles et al., 2019), there is a significant difference in the scale of the different TPSs. A large study of TPSs on web pages (Karaj, Macbeth, Berson & Pujol, 2019) concluded that TPSs owned by Google were present on 82% of the measured web

traffic. This illustrates that Google is operating on almost monopolistic terms and with such a massive access to data, which enables the company to paint a rather revealing portrait of the user.

The meaning of Data Leakage

Literature surrounding TPT, such as Libert (2015) and Krishnamurthy and Wills (2010), frequently refer to tracking as "data leakage", where data is seen as something that "seeps" out of the web page and over to the TPSs.

Some literature has described data leakage as something that happens intentionally or accidentally, such as by the use of malware in applications and where the data leaks in large amounts (Kim, Oh & Kim, 2015). The idea that data leakage occurs as a result of malicious intentions and actions, is also shared by Yu & Guo (2016), as they describe data leakage as a result of poor cyber security measures. This differs from our focus and our definition, as we do not consider TPT as accidental, because the web pages are purposefully requesting these elements. Thus, as it relates to the tracking by TPSs, we define data leakage as *the transmission and flow of user information to TPSs without user consent that happens when a web page is requesting third-party content*. This definition encapsulates the notion that data leakage, in our case, is the result of the purposeful and selective use of third-party content by a web page.

2.3.3 The study of Third-Party Tracking

Philip Agre provides a defining perspective on privacy and surveillance in his article "Surveillance and Capture: Two Models of Privacy" written in 1994. The first model of privacy is that of "surveillance", which follows the assumption that misbehavior is caught and then punished. The second model is that of "capture" which Agre argues is manifested in the practices of information technologists. He further argues that "surveillance" originates from within the classical political sphere as a tool for executing laws, whereas "capture" is more deeply rooted in the practical computer system applications (Agre, 1994).

Capture is understood as "capturing data", which has been enabled by digitalization. Capturing includes telemarketing scripts, toll collection, automated order systems, etc. Agre argued that capture was about "understanding" human activity without awareness of the individual. He believed that both models of privacy were essential in the context of privacy and that surveillance should be seen as a political tool, while capture is more of a linguistic datafied tool. Agre argues that traditional surveillance does not consider the technological development of data collection of the individual, i.e.

higher quantity of data equals higher accuracy (Agre, 1994). The development within tracking has changed tremendously since the paper by Agre. The most recent addition is TPT.

The study of TPT can be traced back to the middle of the 2000's, where Krishnamurthy & Wills (2006) studied prevalence of extraneous content on web pages in the shape of advertisements and the adequate blocking services to avoid these. Few years later, studies found that TPSs collected information about users through the HTTP-requests and cookies. This information was to be considered as personally identifiable information, because it could be connected to other linkable information and thus not only determining the identity of the user but tracking its behaviour throughout the internet (Krishnamurthy & Wills, 2010). This idea that different pieces of information that can all be connected to the same user, is what Libert (2014) highlights as a most problematic aspect of TPT.

The majority of early literature concerning TPT came from scholars in computer science, due to complexity of the topic, which has hindered the possibility for cross-disciplinary work (Libert, 2015). This notion is enforced by Bucy & Zelenkauskaite (2016), who also argue that the technological complexity of big data, which also covers TPT, has made it too difficult to study for social scientific researchers. The ability to access and extract massive amounts of data is still limited to individuals and organizations with the right tools and technical understanding. This has rendered these to become the gatekeepers of the information, as scholars outside of this academic discipline are dependent on those who are technologically skilled for providing research in this field (Bucy & Zelenkauskaite, 2016).

However, both Libert (2015) and Bucy & Zelenkauskaite (2016) argue that the emergence of more easy-to-use tools for data extraction and analysis, could help to bridge the gap between computer science and other academic disciplines. In the case of tracking by TPSs on web pages, this could result in more nuanced research, which could include other aspects than the mechanisms of the tracking itself, such as the social implications.

One such study has been conducted by Helles et al. (2019), who studied the tracking by TPSs across the top 150-websites in the European Union by using the WebXray tool, as presented in Libert (2015). They found that certain trackers only operated or had a strong presence on specific web pages, such as Russian trackers on news portals in Lithuania, Latvia and Estonia, where there is a Russian ethnic minority. This study is a good example of bridging the gap between academic

disciplines on this topic, as it shows that there are distinct relationships between certain web pages and certain TPSs.

2.4 Implications of Tracking

As described in section 2.2, the Danish public sector is showing an increased emphasis on the development of e-government. This entails severe digital changes in the public sector to improve efficiency and free resources. This does however mean that citizens are required to be online, to access the services of the public sector, meaning they are "forced" online. This raises the question of legal implications of public digitalization. We are therefore drawing on the theories by Liebetrau (2017) and Hempling (2014) to better understand how governments are using privately owned IT-infrastructures and what the implications might be as a result. Citizens being "forced" to be online to access public services might compromise the fundamental right to privacy of individuals. We are therefore drawing on theories by Trzaskowski and Sørensen (2019) to explain the fundamental right to privacy and how TPT might compromise this right.

Tracking at the current state is however very uneven, as private organizations primarily are the ones to track, leading to the concept of information asymmetry. We are here drawing on Libert & Nielsen (2018), Thakuriah, Tilahun & Zellner (2017) and Dodds (2017) to offer an understanding of information asymmetry and the implications of organizations being empowered at the expense of individuals. Information asymmetry is further enhanced by information illiteracy also known as digital illiteracy, which is lacking adjustments towards digital technologies, meaning some social groups might be left behind in the digital development of the public sector and other sectors as well.

This section is therefore an account and explanation of why the fundamental right to privacy might be under pressure due to the infrastructural developments within the public sector, resulting in a challenged digital governance. This is further enhanced by the concepts of information asymmetry and digital illiteracy, as organizations are increasingly obtaining more knowledge, while some groups are left behind, making it harder to protect oneself against tracking.

2.4.1 Fundamental Right to Privacy

Privacy is more evident than ever before due to the digitalized global development. The commercialization of surveillance and tracking has led to the privacy of the individual being under pressure due to this extensive collection of personal data. The fundamental right to privacy should

be understood as the right to human dignity and other key values such as freedom of association and freedom of speech and is one of most important rights in the modern digital age (UN, 1948).

In the European Union, the GDPR was recently enforced to ensure data privacy laws of the individual across all member states. Non-compliance with the provisions of the GDPR relating to tracking and data collection of individuals within the union could potentially trigger extensive fines. The purpose of GDPR is to protect individuals of EU member states from the extensive tracking happening at the moment. Organizations outside of the EU collecting data on individuals of the member states are as well subjected to the GDPR. Personal data is defined as "*any information related to a natural person that can be used to directly or indirectly identify that person*" (Kaelin, 2018).

The idea of the privacy concept is to be found in the late 19th century with the publication of "The Right to Privacy" in a Harvard Law Review. The purpose of this publication was to describe how the right to be let alone was a basic human right, with examples as photographs of people in their homes and articles about private life (Trzaskowski & Sørensen, 2019). A later example was the Jewish registry created by Adolf Hitler during the second world war, with detailed information about the location and personal data of all the Jews in Germany, Austria and other occupied countries. The registry was facilitated by Hollerith's, a technology capable of storing data and was produced by the company later known as IBM. The information in the registry led to the death of very large numbers of Jews during the Second World war and was once again an example of the importance of the right to data subject privacy (Trzaskowski & Sørensen, 2019).

In the aftermath of the second world war and to combat the atrocities, the world realized the importance of individual privacy. In 1948, the Universal Declaration of Human Rights was adopted, defining that no one shall be subjected to interference with his privacy, family, etc. The European Convention of Human Rights was shortly after in 1953 adopted, entailing the "Right to Protection of Personal Data". The emergence of information technology resulted in new challenges, as data was increasingly being processed automatically, meaning a detailed safeguard was necessary to ensure privacy. Convention 108 was therefore adopted in 1981 ensuring data subject privacy (Trzaskowski & Sørensen, 2019).

The GDPR is the latest addendum relating to the fundamental right to privacy of the individual and was adopted in May 2018 (Trzaskowski & Sørensen, 2019). The GDPR is a result of the increased tracking caused by the digitalization and datafication and that protection of privacy is more relevant

than ever before. There are however several ways of tracking individuals as previously described. TPT through elements on web pages is for example not covered by the GDPR, which leads to lacking governance and legal frameworks of tracking.

2.4.2 Challenged Digital Public Governance

The public sector has, as previously described, undergone severe digital changes to improve efficiency, leading to the need for competences found in the private sector. The rapid development within information technology affecting social communication and space, as well as governmental affairs, raises the question whether these governmental systems are prepared to deal with the legal implications as a result of the digitalization.

"Law is challenged by the rapid technological development and the complex challenges as a consequence of the digitalization" (Regeringen, 2016)

Organizations are often updating their own standards of data tracking before laws are enacted, to be part of the regulating process and to be part of the standardization of legal frameworks, i.e. private organizations acting as co-regulators (Hempling, 2014). This creates the implication of organizations influencing regulation towards their own interest. It is therefore essential for governments to emphasize impartiality when using private entities as benchmarks for regulation, but also to identify the purpose of the regulation before adhering to industry standards (Liebetrau, 2017; Hempling, 2014). The digitalization is therefore challenging the governmental security of its citizens. The majority of the critically categorized infrastructure of the Danish government is privately owned and run due to stronger competences and more efficient processing. It is essential for the citizens, which live in a state that relies strongly on digital public services, to be ensured security, especially when the majority of the IT-infrastructure is privately owned (Liebetrau, 2017)

In an article by Sten Thorup Kristensen in Advokatsamfundet, he interviews experts on the subject of pitfalls of public digitalization (Kristensen, 2016). A pitfall of the public digitalization strategy according to Jonas Bering Liisberg, CEO at the Danish Parliament's Parliamentary Commissioner, is that it is too focused on user friendliness and not user security. Another pitfall according to Niels Fenger, Professor at Copenhagen University, is that the most vulnerable individuals in a society are challenged even further, as these systems rely on objective information and to a much lesser degree subjective information, which might make it harder for some citizens to grasp the importance of some information and how to fill out specific forms (Kristensen, 2016). The most relevant pitfall of tracking

and the data protection according Rikke Frank Jørgensen, Senior Scientist at Institute for Human Rights is that the Danish Data protection law was based on a directive from 1995, when tracking was based on information citizens were giving in a very limited range of contexts, e.g. at doctors or at social authorities. The data protection does not cover actions in other social spaces, such as on public web pages with TPT. She adds that organizations today also have a stronger interest in possessing online information, compared to when the data protection was put forth, as recent business models are based on data. The last pitfall presented, is that information back in the day was collected according to a “need to have” principle, whereas today is based on a “nice to have” principle (Kristensen, 2016). This is consistent with the concept of surveillance capitalism (Zuboff, 2019)

With the wide adoption of digitalization within public authorities in Denmark, security of the citizens has become highly relevant. In Denmark, the concept of *Digitalization Ready Laws* has therefore become evident with the purpose of ensuring simple and clear legislation, i.e. creating the best public value for the citizens, more user friendly, easily accessible and transparent public sector, while law and order is still enforced (Regeringen, 2016).

However, it is important to remember the conflict between basic rights of individuals and digital law, as legislation has the purpose of regulating behavior, but might also impede the basic rights of the individual. An example would be restriction of freedom of expression, which would undermine democracy. It is therefore evident to discuss the complex and complicated relationship between security and privacy (Duffy, 2019)

The Danish Business Authority launched the Privacy Compass in 2015, which is a tool to aid organizations in handling private data within national and EU law, ensuring compliance within these. The GDPR was also adopted in Denmark in 2018, further improving the protection of the citizens of EU member states (Digitaliseringsstyrelsen, 2016). The Digital strategy of 2016-2020 is discussing 16 specific initiatives to ensure a better inner market in the EU, such as law, initiatives, evaluation of current regulation (Digitaliseringsstyrelsen, 2016)

The question is however, if legislation will become more flexible in the context of an ever-changing digital landscape, as some forms of tracking, i.e. TPT is not covered by existing laws, hereby undermining the fundamental right to protection of privacy as previously stated.

2.4.3 Information Asymmetry and Digital Illiteracy

Data is shaping the way we live; hence it is called the Digital Age (Plesner & Husted, 2019). Data is affecting organizations, states, governments, citizens, and more. We have therefore entered the data economy, where search and transaction cost are reduced and data could be used as a facilitating tool within medical and scientific research, which in the end leads to higher efficiency. Data is even used within public organs to deliver more efficient services and better social welfare (UN, 2019).

The amount of data produced and left behind by users online is vast in size, as 6,2 billion Google searches are conducted each day, as 238,7 billion emails are sent every day, etc. We are producing 5,8 billion gigabytes of data each day at the current rate, meaning enormous amounts of data to be used for tracking purposes. These enormous amounts of data are however dispersed over the entire web and are very messy, meaning complex methods of data processing is needed for the data to offer insights (UN, 2019).

The data economy is currently very uneven, equaling asymmetric information power between organization-organization, organization-consumer and organization-state relationships. The unique property of data makes these asymmetries very hard to reduce, as data depends on very complex infrastructure and monopolistic positions to gain the best possible insights, as is the case with Facebook, Google, Microsoft etc. Smaller firms in the field of data competition have no way to compete with the established firms due to the complexity. From a political economy perspective, these concentrations of power might increase the chance of "regulatory capture", which is "*a situation where policymakers or enforcement agencies are in a constant state of being influenced by powerful firms*" (Hempling, 2014:5). The power of these organizations in both the economic and political perspective might therefore impede the freedom of the user and be a barrier for democracy.

The initial private organizational data tracking was based on an advertising incentive, i.e. targeting specific individuals with specific products. The business model this was based on, was as previously described based on a broader social agreement between the public and organizations of a monetarily free product in exchange for being exposed to advertising. This development within online behavioural advertising (OBA) has allowed for systems capable of covertly surveying consumer preferences. Due to lacking regulation within this arena, internet users have no control over this surveillance and could again be impeding freedom of choice and democracy (Libert & Nielsen, 2018).

OBA represents an extensive shift in the balance within society, which is caused by the concept *information asymmetry*. Information asymmetry is defined as a "*situation that favors the more knowledgeable party in a transaction*" according to the business dictionary (BD, 2020). An example of information asymmetry is the access to the free to use service Google Maps. This service has become very valuable for individuals as it offers detailed location mapping. It is important to emphasize the trade-off between the individual and Google, as Google gains tremendous insights into very specific information such as location of parking spaces, most popular parking spaces and roads, availability of parking spaces, etc. This allows google to offer very specific services for its users due to aggregation through algorithms and big data. This transaction is illustrated in the figure below.

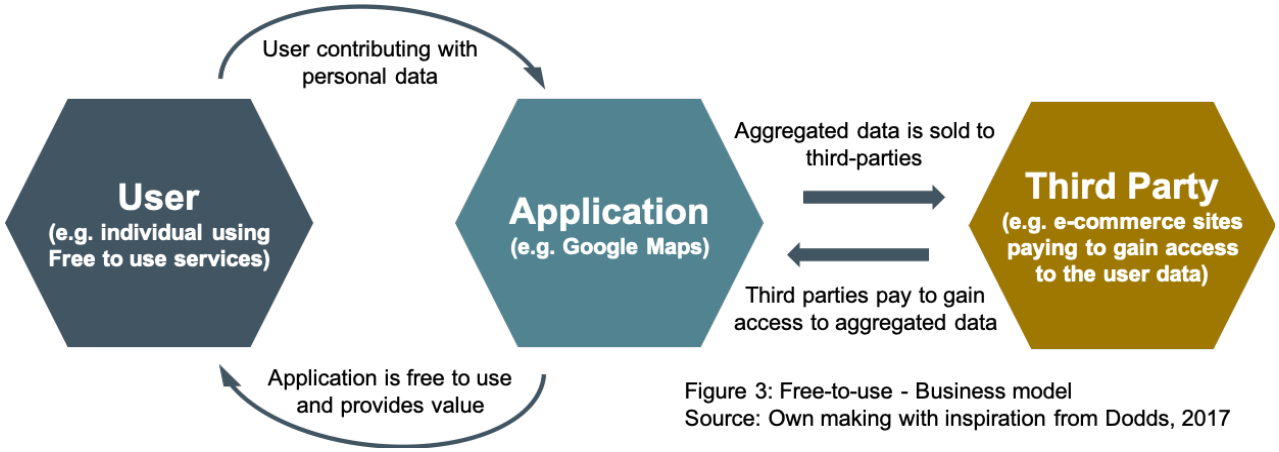


Figure 3: Free-to-use - Business model
Source: Own making with inspiration from Dodds, 2017

Data asymmetry and the resulting imbalance of power is often raised in the context of personal data according to Dodds (2017) and Libert & Nielsen (2018). Organizations are empowered at the expense of the users. This imbalance is giving organizations and governments increased knowledge and hereby power over citizens. The current situation is therefore organizations knowing more about citizens than citizens possessing information about organizations hence the information asymmetry (Libert & Nielsen, 2018).

The information asymmetry is further enhanced by the concept *information illiteracy* also called digital illiteracy. Literacy is defined by the OECD "*as the ability to understand and use information to expand one's capacity*" (Margerie, 2018). Digital illiteracy on the other hand is defined as "*Digital gap emerged in certain social groups that have been left out of this technological process derived from their maladjustment to new technological developments*" (IGI Global, 2020). In the context of

digital illiteracy, the OECD found in a study from 2018 that 70% of the population between 16 and 65 lack expertise within basic computer science (Margerie, 2018).

Digital illiteracy has become more evident than ever before, as the world is increasingly becoming digitalized. Digital illiteracy has created the digital divide, also called the “participatory gap”, which states that even individuals with computers, smartphones, internet, etc., lack the skills, education and familiarity to leverage the opportunities of the digital age. The differences in digital literacy have been correlated to patterns of social exclusion in society according to Warren (2007), Lee et al., (2015) and Mossberger et al., (2012) in Thakuriah et al., (2017). The current social divide is not a result of the generation divide, it is however a result of socioeconomic status (Thakuriah et al., 2017)

With the general adoption of the concept e-government, we are witnessing how public sectors are increasingly digitalized and hereby reliant on digital technology, navigate public web pages, electronic documentation and other important digital skills in the public sector. The participatory gap or digital illiteracy will undoubtedly increase complexity for individuals, organizations, etc., who are forced to access these public web pages (Thakuriah et al., 2017).

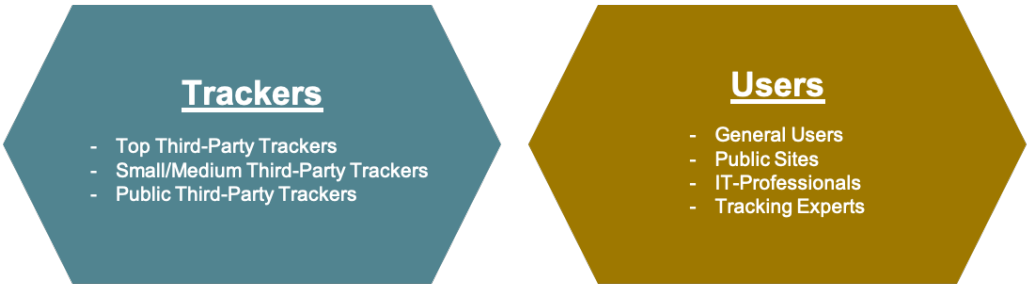
2.5 Stakeholder Theory

The purpose of our stakeholder analysis is to provide an overview of the different stakeholders relevant to the TPT on Danish public web pages and to provide an overview of the relations and interdependencies between the different stakeholders. There are several different definitions of a stakeholder, such as the definition by Freeman (1984) stating that stakeholders are individuals or groups who affect or are affected by an organization’s objectives. Classical stakeholder theory usually applies a neoclassical private organizational perspective, viewing the stakeholder from a specific organization’s point of view (Mitchel, Agle & Wood, 1997). The point of view in our study is more of a general view. We are going to be evaluating the stakeholders based on whether they contribute negatively, positively or neutrally to the ecosystem as, e.g. TPSs associated with malware would be seen contributing negatively, whereas experts of tracking would be contributing more positively.

We therefore draw on Scholl (1970) as the article provides an overview of applying stakeholder theory to public sectors and hereunder the e-government. The public sector of today is more of a multi-jurisdictional and multi-sector endeavor, meaning that public-private collaboration could lead to different governance issues (Scholl, 1970). We therefore draw on a combination of Mitchell et al.,

(1997)'s concept of stakeholder identification, i.e. power, legitimacy and urgency and on Blair & Whitehead's (1988) diagnostic topology of a stakeholder's potential for collaboration vs. potential for threat. This theoretical combination is according to Scholl (1970) an optimal solution for public sector stakeholder analysis. Power should be understood as the power of influencing towards an objective and according to Mitchell et al., (1997) allows for three different kinds of power, i.e. political power, resource power, and normative power. Legitimacy should be understood as actions being appropriate within the accepted system. Legitimacy in our case means actions being appropriate to the protection of the individual. Urgency should be understood as the urgency for attention, i.e. time sensitivity and criticality (Mitchell et al., 1997). The diagnostic topology also takes relative power and resources into consideration and the stakeholder's ability to either threaten or cooperate.

We are therefore drawing on the theories described above, hereby looking at the different stakeholders of the TPT ecosystem through the above described attributes and in relation to each other. We are going to be analyzing the following stakeholders, as these are the once we found interesting throughout our analysis:



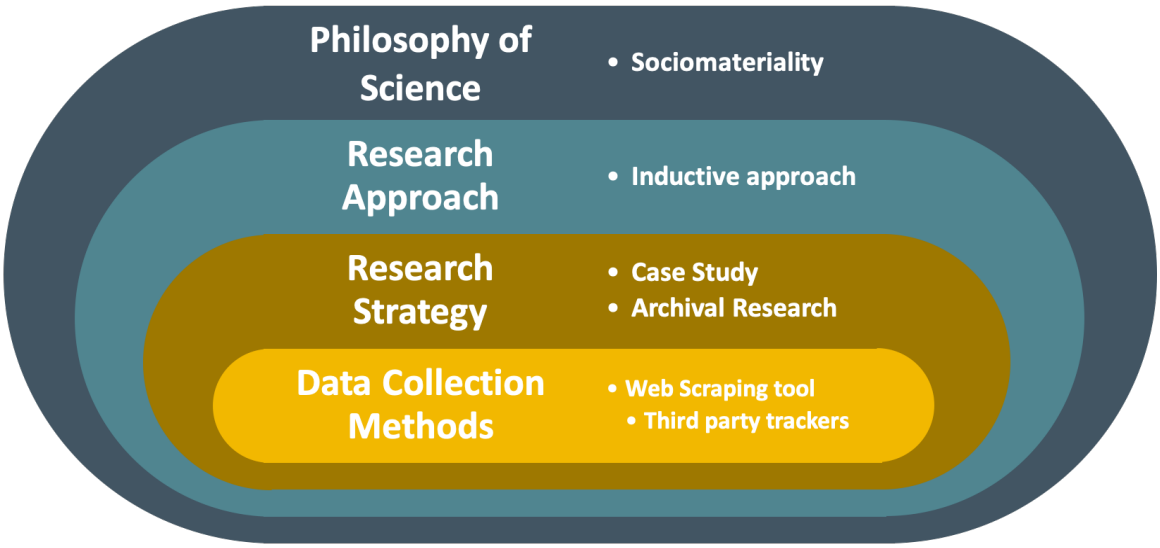
Source: Own making

3. Research Methodology

"...nearly nine in ten websites leak user data to parties of which the user is likely unaware of … In order to detect tracking (third party tracking) on the sites selected, the WebXray software (version 1.0) was used. Third-party HTTP requests provide an excellent unit of analysis for examining the extent of tracking on the web. Such requests are potentially as revealing and invasive as cookies…"

Tim Libert (Libert, 2015:51)

This section is a description and explanation of our methodological approach. Our methodological approach has been structured according to the principles of the *Research Onion Diagram* (Saunders, Lewis & Thornhill, 2009). We have chosen the relevant parts of the diagram according to our study, i.e. philosophical point of departure, research approach, research strategy and choice of methods and data collection methods. We have illustrated our approach in the figure below, starting with the outer layer of Sociomateriality.



Source: Own making with inspiration from Saunders et al, 2009)

According to Werneck (2006) in Dresch, Lacerda & Antunes (2015), knowledge production can be understood as the "*construction of universally accepted knowledge in a given historical time or as a process of learning of the subject*" (Weneck, 2006, as cited in Dresch et al., 2015:175).

In our thesis, we are using the factual science approach, which has the purpose of exploring, describing, explaining and predicting compared to formal science, which does not depend on an empirical basis. Our thesis is based on empirical data retrieved from web page analysis.

Factual science is divided into social- and natural science. Our approach is a combination of social science and natural science hence our philosophical point of departure being that of sociomateriality. Sociomateriality allows for looking at natural science by identifying and understanding the complex phenomena in its raw state and to discover why things are as they are within data. The primary purpose of using natural science is to discover and justify the reason for our empirical findings related to the observable facts. Sociomateriality also allows for looking at social science, which seeks to describe, understand and reflect on human beings in the equation. When drawing on social science,

we are able to identify how different actors are present in the context of data and tracking. Social science is based on subjectivity and should therefore be used with critical emphasis (Dresch et al., 2015; Yang, 2016).

3.1 Philosophy of Science

The following section is our philosophical point of departure. We are here stating how, within the philosophical discipline of Sociomateriality, knowledge is acquired and how this philosophy is going to aid us in acquiring knowledge.

3.1.1 Sociomateriality

Sociomateriality is one of the more recent paradigms to have surfaced within the philosophy of science. It is a growing research stream within the field of information systems and information communication technology (Yang, 2016).

With the evolution of the internet and the rapid expansion of IT-infrastructures, e-commercialization, the digitalization of business models, etc., changing market- and social dynamics are a fact. We are therefore witnessing defining developments both within the business arena, but also within the social arena, as the line between these arenas is getting increasingly blurry. The way we communicate as humans has undergone significant changes due to the changing material base on which the global information and communication infrastructure is built (Ciborra, 2006). The changing dynamics and understanding of the world has therefore become increasingly complex and literature within this field, such as Elovaara, 2004; Orlikowski, 2007; Schwanen, 2008; Orlikowski and Scott, 2008; Scott and Orlikowski, 2009, argue that the ontological approach to understand these changing dynamics is the be found using sociomateriality in Yang (2016).

We therefore apply sociomateriality as a point of departure to understand the world as it is. This allows us to combine the ontological fusion of technologies as a material artefact and humans and institutions as actors and interpreters. Sociomateriality is therefore a compromise between the polarized field of technological determinism and social constructivism. Technological determinism is where artefacts are seen as limited entities with intrinsic features that cannot be denied and relies on the causal relationship between technology and human activity, i.e. technology deciding human action. On the other hand, social constructivism is where everything is articulated, perceived and interpreted by humans and is therefore the discursive representation of technology and how it affects human activity (Plesner & Husted, 2019).

By viewing the world through the lens of sociomateriality, we are able to understand technology in the context of actors influencing and being influenced by other actors. This allows us to gain better insights and to identify the role of technology in shaping networks of different actors, i.e. understanding the material properties, the subjective interpretations and the social context all together (Yang, 2016).

In the context of our study, this allows us to better understand data as an entity with built in technological features, to understand how people perceive data through discursive interpretation, but also how the combination of the technological features and discursive interpretations allow for different and polarized understandings of data as a technology. By applying sociomateriality, we are able to better understand the different purposes of tracking and the different uses of data and it is therefore relevant in the case of TPT on web pages.

3.2 Research Purpose & Goal

According to Booth et al., (2008), scientific research can be based on one or more reasons for conducting the study such as new and interesting information, an answer to an important issue or an in-depth understanding of some phenomenon. The approach of our thesis is to provide an answer and understanding of an important issue, i.e. an understanding of the TPT on the Danish public web pages, hereby providing insights into how to approach this issue (Dresch et al., 2015; Saunders et al., 2012).

Our motivation for this study was based on the study by Helles et al. (2019) who wrote the article “Infrastructures of tracking: Mapping the Ecology of TPSs Across Top Sites in the EU”, which is the study of third-party trackers in the EU. We found this article particularly interesting due to the fact that TPT is not covered by current regulation and might therefore be a breach of consent and privacy and due to the extent of data leakage the study found. This thesis is, compared to existing literature, studying TPT with a focus on Danish public web pages. We chose the focus on Danish public web pages, as Denmark wishes to be a frontrunner in terms of e-government, forcing the citizens to leverage electronic means of public service. The motivation of our thesis was therefore based on a gap in existing literature, as we are studying a specific part of the ecosystem, the part where personal information might be more vulnerable and with little to no protection (Dresch et al., 2015; Helles et al., 2019).

The goal of our study is to explore, describe and explain the phenomenon of TPT on Danish public web pages, hence our problem statement: How can we understand the extent of TPT on Danish public web pages and the underlying ecosystem of third-party trackers?

3.3 Research Approach & Strategy

After having covered the purpose, strategy and goal of our study, this section will cover the main scientific methods guiding our study.

3.3.1 Scientific Approach: Inductive

In our study of TPT on Danish public web pages, we are using the inductive approach. This approach is our point of departure, from which we are studying this phenomenon, as we are studying empirical observations. The purpose of the inductive approach in our study is to identify correlations and patterns as well as find possible explanations for these etc., within the data extracted. We wish to be able to say something theoretical about Danish public web pages in terms of TPT. The strategy is therefore analyzing something specific, to be able to say something general about the landscape (Andersen, Hansen & Klemmensen, 2012).

Figure 4: Inductive approach
Source: Own making with inspiration from Andersen et al, 2012

As the above figure illustrates, we are going to start from the assumption that we will be able to understand the extent of TPT and the underlying ecosystem within Danish public web pages by observing the extracted data from WebXray (WebXray will be covered in section 3.5). The understanding of the ecosystem is based on an analysis of patterns and correlations within the ecosystem, which in the end allows us to understand and say something general about TPT within Danish public web pages and the underlying ecosystem (Dresch et al., 2015).

The inductive approach has been criticized for “the inductive leap”, which is a critique of the generalization based on a single phenomenon. We do therefore not expect to provide an in-depth

generalized theory of TPT within Danish public web pages, but expect to say something general about the extent of tracking, to map the ecosystem and for future research (Saunders et al., 2012)

3.3.2 Strategic Research Method: Case Study

Having elaborated our scientific method, we are in this section addressing the specific research method of our thesis. The research method of our study is a case study of the extent of TPT on Danish public web pages and the underlying ecosystem of third-party trackers.

The purpose of a case study according to Yin (2013) is to investigate an empirical phenomenon to achieve a better understanding of this usually complex phenomenon in its real context. In the context of our study, we are going to investigate TPT within Danish public web pages to understand this phenomenon in its real context. By using the case study as a research method, this allows us to provide a detailed description of the phenomenon.

The case study allows for an in-depth understanding of the phenomenon, as case studies usually consist of several factors, including, but not limited to identifying and analyzing the ecosystem, regulation, stakeholders, data leakage, usage vs. abusage, risks and consequences of large scale data collection, privacy and consent in the context of each other (Dubé & Paré, 2003; Eisenhardt, 1989). According to Eisenhardt (1989) the foundation of a case study is contextualizing collected data in its real context, where the researcher seeks to identify theoretical categories to be used for the proposal of new theories. We are therefore, as previously stated from the figure above, moving from empirical findings to a more generalized theoretical ground with the “inductive leap” taken into consideration.

The inductive nature of our study is a result of the case study, as our study’s point of departure is based on empirical findings about TPT on Danish public web pages, which is a real-life context. Another defining factor for the research method of choice, is our position as researchers, as we are not interfering with the process of data collection, but rather acting solely as observers and collectors.

According to Ellram (1996), case studies are furthermore exploratory, descriptive and explanatory, which is typical for natural and social sciences and allows for the identification of behavioral patterns to explain the phenomenon. The theoretical framework surrounding the case study therefore allows us as researchers to explore, describe and explain the complex phenomenon of TPT on Danish

public web pages by identifying and analyzing behavioral patterns within the ecosystem and hereby contextualizing these patterns into our conceptual framework.

3.4 Data Description

The dataset for our thesis consists of information about TPSs found on web pages belonging to the public administration. These web pages are defined in the Danish government's official document regarding the structure of the public administration (Digitaliseringsstyrelsen, 2019). Though the document concerns the Kingdom of Denmark, we have decided to exclude the web pages belonging to the public administration in Greenland and Faroese Islands, hereby focusing solely on the web pages belonging to the public administration in Denmark.

Furthermore, web pages of the Danish public hospitals and psychiatric departments have been added to our list of web pages that we analyze. Only one region, Region Hovedstaden, has distinct web pages for all their hospitals. Three other regions have clustered two or more nearby hospitals together into a hospital entity, where these entities have their own distinct web pages. Only one region, Region Sjælland, does not have distinct web pages for any of their hospitals and thus no distinct web pages for the hospitals nor psychiatric departments from Region Sjælland have been included in the web page list. All the regions, except Region Sjælland, have distinctive web pages for their psychiatric departments, which have been included as well.

Combining the selected web pages from the official document (Digitaliseringsstyrelsen 2019) and the web pages of hospitals and psychiatric departments of the different regions, the number of web pages to be analyzed are 285.

Types of Data

In order to study the tracking on web pages by TPSs, we analyze the HTTP requests, because this mechanism can transfer user data to TPSs. This method of studying can help us to understand the extent of TPT on Danish public web pages, as well as how the web page uses TPSs and how this usage may cause data leakage to TPSs.

The study is conducted using a mixed methods approach, where the quantitative data is retrieved by using the WebXray tool to collect and analyze the HTTP requests. Through the analysis of the ecosystems, the quantitative data will be converted into qualitative data, by identifying the specific relationships between the web pages of the TPSs. We also include secondary data to analyze these relationships. According to Zeller (2017), web tracking can be characterized as a method that draws

from the mixed methods approach. However, Zeller (2017) defines web tracking as the tracking of users on web pages, in order to study the user, whereas we seek to study the TPT of users. The ethical aspects of this difference will be further outlined and covered in section 3.8 Ethical Considerations.

The secondary data in this project has primarily been used for conceptualizing the scene for why this phenomenon needs to be studied. This data has been retrieved using a bibliographic technique, where the researcher examines existing literature in order to find any gaps or areas that need further exploration and detailed studying (Saunders et al., 2012). One such example is from Helles et al. (2019), who propose that the public administered web pages should reconsider their use of TPSs, as they are not lying in competition with other web pages.

3.5 Data Collection Methods and Tools

For analyzing the HTTP requests, we use a software called WebXray. WebXray is a tool coded in python by Tim Libert and was first presented in Libert (2015). The empirical output that we collect by using WebXray serves as our primary data. The WebXray software is an example of the bridging between computer science and other academic disciplines (Libert, 2015), as it is open source, easily accessible and easy to use. WebXray has been used in other scientific studies of TPT on web pages, including Libert (2015) and Helles et al. (2019).

In order to retrieve data about the HTTP requests, WebXray works on multiple levels. Firstly, a list of the URL-addresses of the selected web pages in standard text file format, must be loaded into the program. WebXray then uses the Google Chrome web browser to load and run the pages one at a time, in order to analyze each page for third-party cookies and elements. When analyzing a page, WebXray registers all the HTTP requests that are made from the specific web page (Libert, 2015). The process is then repeated on all the listed pages.

When analyzing the HTTP requests, WebXray looks for indicators that can define the type of requested element or cookie, as well as information about the involved TPS. Some web pages use their own servers, where the requested content is located. This means that user data does not leak to a TPS, even though the web page receives the desired content (Libert, 2015). Essentially, if the domain of the web page, *e.g. "http://masterthesis.com"* is found in the request string for the element, e.g. "*http://images.masterthesis.com/contact.jpeg"*, it means that it is not a third-party request. In our analysis, this kind of HTTP request is not considered as a third-party request and thus does not

contribute to the leakage of user data. For that reason, these requests to own servers are excluded in the analysis.

Even though HTTP requests to own servers do not contribute to leakage of user data, the requests for third-party content do. HTTP requests to TPSs can also be identified from their request string. WebXray analyzes the request string by dissecting it based on indicators, which the following hypothetical request string exemplifies:

http://sub.domain.com/picture.png

This request string shows the domain "*domain.com*" and the subdomain "*sub*". It also shows the name of the element "*picture*", however WebXray uses the file extension "*png*" to determine the type of element. Since "*png*" (portable graphics network) is a file extension for images, WebXray can determine that this element is an image. Other types of elements with their own file extensions, such as JS (JavaScript) and CSS (Cascading Style Sheets) can also occur. WebXray can also couple certain domains to ownership structure and country of origin. This is not information that can be found in the request string, but it is found through a WHOIS search, which is a web page domain lookup tool. In cases where the WHOIS search result is insufficient, the information has been found through web searches and thereafter coded into WebXray by the developers (Libert, 2015). The ownership structure is relevant to include, because some companies that provide TPSs have been found to use different domains. One such example is Adobe that owns the domain "*adobe.com*", but also, and perhaps less obvious, owns the domain "*2o7.net*" (Libert, 2015). This is very important, as even though these domains are different, the user data will still be transmitted to the same company or organization. The country of origin of the TPS is also relevant, as it helps to map the funneling and transmission of user data. Furthermore, depending on the country, the legislation for handling this user data by the TPSs could differ.

The information that WebXray gathers from the scan and analysis is then exported to an SQLite database file. SQLite is a type of relational database system that uses Structured Query Language (SQL) to handle the data. This type of data handling is ideal for this study, as it allows the researcher to identify relationships between the data points across different tables. The SQL queries can be found in appendix A and B.

3.6 Sample Selection

In the data collected from WebXray, we used the tables "*cookie*", "*element*", "*domain*", "*domain_owner*" and "page", since only these tables contained relevant information to our study.

However, within each of these tables, only certain columns were relevant to include in our analysis. The excluded tables are called "*error*", which lists the pages that WebXray failed to load, and "*sqlite_sequence*", which lists the number of records in each table. The full unedited SQLite file including all tables can be found in appendix C within a link to GitHub.

The included columns within each of the relevant tables are mainly primary keys, but certain foreign keys were also included. The foreign keys were included in order to be able to join one table with another table containing the respective primary key. A description of primary and foreign keys can be found in section 3.7 Data Transformation.

Cookie Table

We used four columns from the table "*cookie*", as seen below.

Column	Primary or secondary key	Description
id	Primary key	Unique identifier for the cookie request
page_id	Foreign key from ”id” in ”page”	Unique identifier for the page
name	N/A	Name of the cookie
domain	N/A	Name of the third-party domain
domain_id	Foreign key from ”id” in ”domain”	Unique identifier for the domain

Table 1: Cookie table

These columns contain the most relevant information about the cookies. Compared to elements, these cookies have fewer distinguishable characteristics. The characteristics of the cookies will be covered in the analysis.

Element Table

From the table "*element*" we included five columns, as listed below.

Column	Primary or secondary key	Description
id	Primary key	Unique identifier for the element request
page_id	Foreign key from "id" in "pages"	Unique identifier for the page
element_url	N/A	Contains the name of the element
domain_id	Foreign key from "id" in "domain"	Unique identifier for the domain
extension	N/A	Indicates files extension of the element

Table 2: Element table

We included the column "*element_url*", since this column contains the file name of an element and thus serves as a unique identifier for this specific element, such as "*https://abs.twimg.com/emoji/v2/72x72/2708.png*". We also included the column "*extension*",

because this includes the file extension of the element, which can determine the type of element, such as the image file extension "*png*". The columns "*content_type*" and "*type*" also indicate the type of element, however these are not as specific as "*extension*", and thus were excluded.
The foreign key "*page_id*" was also included, which is a unique identifier for the web pages, or *pages*, in the pagelist and is native to the table "*page*". The column "*is_3p*" was not included in the results like the other aforementioned columns, however it was used to filter out non-third-party elements.

Domain Table

Four columns were included from the table "*domain*", as seen below.

Column	Primary or secondary key	Description
Id	Primary key	Unique identifier for the element request
domain	N/A	Name of the domain
tld	N/A	Indicates top-level domain
domain_owner_id	Foreign key from "id" in "domain_owner"	Unique identifier for the domain owner

Table 3: Domain table

The column "*domain*" is included, because this contains the domains that host third-party content, such as "*twimg.com*". The column "*tld*" was also included, as it contains the top-level domain of the third-party domain, such as "*.com*". The column "*Pubsuffix*" is identical to "tld" and using either would have worked. The column "*name*" was also included, since it indicates the company, program or organization that the third-party domain belongs to.

Domain_owner Table

From the table "*domain_owner*", we included three columns, as listed below.

Column	Primary or secondary key	Description
id	Primary key	Unique identifier for domain owner
name	N/A	Indicates owner of third-party domain
country	N/A	Indicates country of origin of third-party domain owner

Table 4: Domain table

The column "*country*" indicates the nationality of the third-party domain, such as "*US*", which indicates the United States of America. The columns "*name*" and "*country*" indicate the owner company and its country of origin.

Page Table

From the table "*page*", we included two columns, as listed below.

Column	Primary or secondary key	Description
id	Primary key for page	Unique identifier for the page
start_url	N/A	Indicates the URL of the page

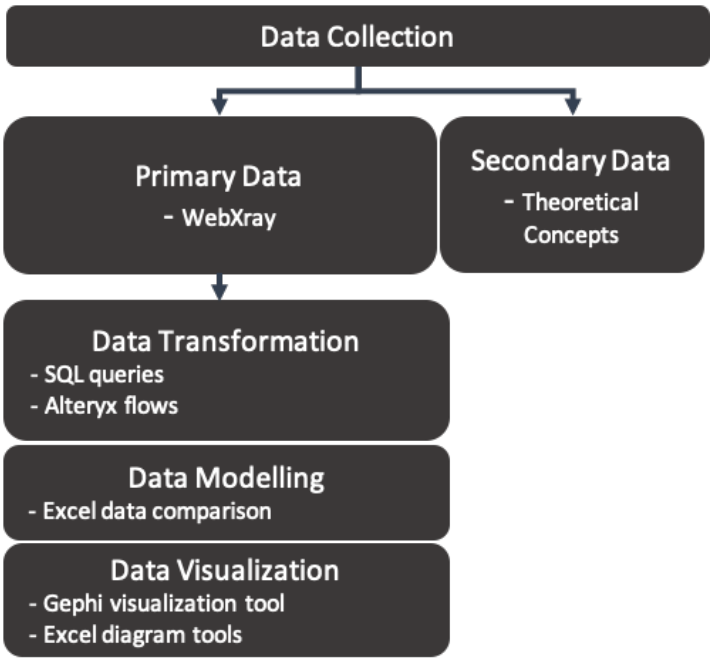
Table 5: Page table

The column "*id*" is the unique identifier for the page and the columns labelled as "*page_id*" in the other tables are based on this column. The column "*start_url*" indicates the url of the web page that is loaded in the browser, before any possible redirection, such as "*http://www.um.dk*".

3.7 Data Transformation, Analysis & Visualization

WebXray was, as previously stated, used to collect our data. The output of the data collection was then transformed using two different tools. Alteryx was initially used to clean the data to get a better overview of the entirety of the data. Alteryx is a business intelligence software tool for cleaning, analyzing, transforming and visualizing data (Alteryx, 2020). The cleaned data from the flows were not used for any analysis, but just to get an overview of the structure and totality of the data.

SQL queries were however used more extensively to retrieve the necessary information of the database to be used for modelling and visualization of the data. In order to analyze the third-party element requests, the data output from WebXray was filtered in a SQLite relational database. Firstly, we combined the relevant tables, which contained all of the specific data pieces. We then filtered the sample selection, to only include the elements labelled as "third-party", which were found using WebXray's own third-party filter called "*is_3p*".



However, some elements were not caught by the filter, because they shared their domain name with the pages that they were found on. To solve this issue, we manually checked each element to identify if it appeared on any other pages than the one it shared domain with. If not, it was excluded. One

such example is the domain "*esbjerg.dk*". The domain "*esbjerg.dk*" received requests from 27 different elements, all of which were requested by one page, namely "http://www.esbjerg.dk". We therefore excluded elements from the domain "*esbjerg.dk*", as it is not a third-party domain.

In the processing of the WebXray results for cookies, we included all results, regardless if the cookies came from third-party domains or not. We made this decision due to the fact that cookies, compared to elements, are able to collect more data about the user, such as how the user behaves in the browsing session (Libert, 2015). It is our understanding that this decision will help to create a better understanding of the usage of cookies on Danish public web pages and to ultimately help to better understand TPT on Danish public web pages. We did not accept any of the cookies prior or during the WebXray analysis, which means that all the cookies found by WebXray, were set without consenting.

Out of the WebXray output, see section 3.6 Sample Selection for a description of the output, only certain tables and columns were relevant to include in the analysis and therefore, tables containing relevant columns were combined successfully in the SQL query to allow for further processing. This is due to the fact that some tables contain some of the same values, e.g. one table contains a foreign key column, which is based on a primary key column in the other table. A *primary key* is a column that works as a unique identifier (ID) for that specific table and originates from this table. When a *primary key* is included in any table other than the one that it is originally from, it is called a *secondary key*. In other words, the *primary key* is the original column and the *secondary key* is the copy or duplicate in another table.

For example, the table "*element*" could be joined with the table "*domain*", because the values in the column "*domain_id*" (Foreign key) in "*element*" is identical to values in the column "*id*" (Primary key) in "*domain*". Thus, the SQLite database will know that values in rows with "*id*" = X should also be attributed to rows with "*domain.id*" = X. This meant that query results with more characteristics could be achieved, such as retrieving element URLs, a domain and the nationality of the domain, since the "*country*" column, which denotes nationality, is only found in the "*domain*" table. After having used SQL to query the specified data, we used Excel combined with the SQL output to model the data and get a better understanding of which cookies and elements were present on which web pages. This was our initial relation-analysis, allowing identification of the extent of cookies and elements and the relationship between the different trackers and the different web pages. The analysis of the relationships through SQL and Excel resulted in our initial ecosystem analysis.

The relationships between the trackers and pages were however difficult to identify and analyze manually in Excel, we therefore used Gephi as a visual analytics tool to visualize our ecosystem. Gephi is a visualization and exploration software for graphs and networks (Gephi, 2020). By using Gephi, we were able to through exploratory- and link analysis to reveal the underlying structures of associations between the trackers and web pages in our ecosystem, i.e. identifying how the different trackers and pages were interconnected. Gephi relies on input based on *nodes*, which are the actors, i.e. the trackers and web pages, and *edges*, which is the relationship between the nodes, i.e. where the trackers are present on specific pages.

We here used the modelling from our Excel sheet and transformed the data once more by providing each tracker with a unique ID and each page with a unique ID and we were hereby able to identify on which unique pages the different trackers were present. For example, ID 293 was present on ID 5, 98 and 179, i.e. three different web pages. For Gephi to be able to understand the inputs and the relations between the inputs, the trackers and web pages were given IDs within the same list, meaning that the web pages were given IDs from 1-277 and third-party domains that set cookies were given IDs from 278-329. For the element tracking ecosystem, web pages were given IDs ranging from 1-277 (the same as in cookie ecosystem) and the tracking elements were given 278-393. This allowed us to visualize trackers and their relation to web pages within the ecosystem, which will be presented in our findings and analysis section. We produced two different ecosystems, one for cookies and one for elements.

Just uploading the nodes and edges to Gephi did however not show anything relevant, as all of the nodes and edges were just one big chunk of black dots. Within Gephi we therefore transformed the visuals of our ecosystem by setting different parameters. The first thing was changing the size of the nodes in relation to the number of outer edges, meaning that trackers who were present on many pages were proportionally bigger than trackers with less presence. After changing the size, we applied colours relating to the modularity of the trackers, meaning that we applied colour related to the clustering of the trackers. Trackers with more than one relation to each other, i.e. tracking the same web pages, were therefore part of the same cluster and therefore had the same colour.

We did however find that some of the trackers were present on a great number of web pages, such as third-party domains from Google Analytics and Siteimprove. We therefore split the cookie ecosystem into two separate ecosystems and split the element ecosystem into two separate ecosystems. One ecosystem containing the trackers with a presence on 10+ web pages and one

where trackers with a presence on 10+ web pages were removed, meaning that we ended up with four ecosystems in total. We included both ecosystems for respectively cookies and elements to show the complexity of the ecosystems with the top trackers, but also included the ones without the top trackers, in order to be able to identify the specific and underlying clusters within the ecosystems.

The visualizations within our thesis serve different purposes. Some of our visualizations serve the sole purpose of making it easier for the reader to get an overview, such as the visualization used in the research question section. This visualization could have been left out, and the result would have been the same. According to Kennedy & Engebretsen (2020:19) "*visual representations of statistics and other, often quantitative data can convey complex facts and patterns quickly and effectively*".

Other visualizations within our thesis serve a more specific purpose relating to the above quote, such as the *information cycle* in section 2.1.2. This illustration would be very hard to describe for a reader, as it contains many different factors. We therefore chose to illustrate and then explain the process below, hereby leading the reader through the cycle. Another example is our visualizations of the ecosystems in our findings and analysis section. The data behind the ecosystem as previously described does not illustrate anything without having existing knowledge. We therefore visualized the ecosystems to show the complexity and relations between different trackers and web pages, which would not have been possible without an illustration. Our visuals therefore serve the purpose of benefiting the process of sense-making and learning through the thesis and are tools for understanding (Kennedy & Engebretsen, 2020).

3.8 Ethical Considerations

Our data collection method is defined as internet research according to the AoIR Ethics Working Committee, as the application of WebXray is categorized as data scraping. It is further categorized as scraping and data storage due to how the data is utilized and processed, which in our case is through SQL, Excel and Gephi. The purpose of the data processing is to ultimately employ visual-, descriptive-, exploratory- and link analysis (Markham & Buchanan, 2012).

The ethical considerations of our study are however deviating from regular data scraping, as the data we are collecting is publicly available, but the right tools are required to gain access, hence the use of WebXray. The data we are collecting does not disclose any personal identifiers but does disclose organizational identifiers. These organizational identifiers are web page addresses, meaning they are available for all to see. We are not collecting personal data, nor are we collecting

harmful data about organizations. We are however collecting data about the third-party elements and cookies that can be used for tracking (Markham & Buchanan, 2012). The initial Internet Research Ethics guidelines, later called IRE 1.0, presented by Ess and the AoIR Ethics Working Committee (2002) was a first step towards ethical internet research, ensuring proper protection of the different actors in the research process. In 2012, IRE 2.0 was published and worked in conjunction with IRE 1.0 to offer more clear and processual ethical guidelines for conducting internet research, as the internet in the period was undergoing significant changes (Markham & Buchanan, 2012).

The key guiding principles of ethical studies put forth by the AoIR Ethics Working Committee in IRE 2.0 state that the greater the vulnerability of the participant (organizations and human subjects in our case), the greater the obligation to protect the participant from harm. Harm should be understood contextually and not universally, meaning that ethical principles should be understood in an inductive manner, rather than being the same for all cases (Markham & Buchanan, 2012). Ethics should therefore be understood as a process approach in the context of case-by-case situations and in relation to the methodological considerations of the study. Our ethical considerations are therefore based on the balance between the rights of organizations tracking through TPSs, the public web pages and the rights of data subjects (human subjects) who are being tracked, but also the social benefits of the researcher's right to conduct this research.

The concept of *human subject* has for an extended period of time been seen in relation to medical experiments and interview and/or questionnaire studies. Within IRE, human subjects should also be understood as subjects who are studied without their consent, e.g. through TPT, as tracking might lead to harm, vulnerability and personally identifiable information. The varied understanding of the human subject is also emphasized in the IRE 3.0 guidelines. The IRE 3.0 guidelines are cited as Franzke, Bechmann, Zimmer, Ess and the Association of Internet Researchers (2020).

Our primary focus is not researching *the system*, i.e. how and why organizations are tracking and being tracked. This is a by-product of the primary focus, which is how the ecosystem is currently organized to uncover possible harm for human subjects who are using Danish public web pages. This does however mean that Danish public web pages with TPSs might be subjected to negative public display, which could cause unfavorable publicity and reputation. By undertaking a human-subject-harm-focus as we do in this thesis and balancing it towards a system-harm-focus, we are arguing that the ethical balance is in favour of conducting this study. Personal data leakage from

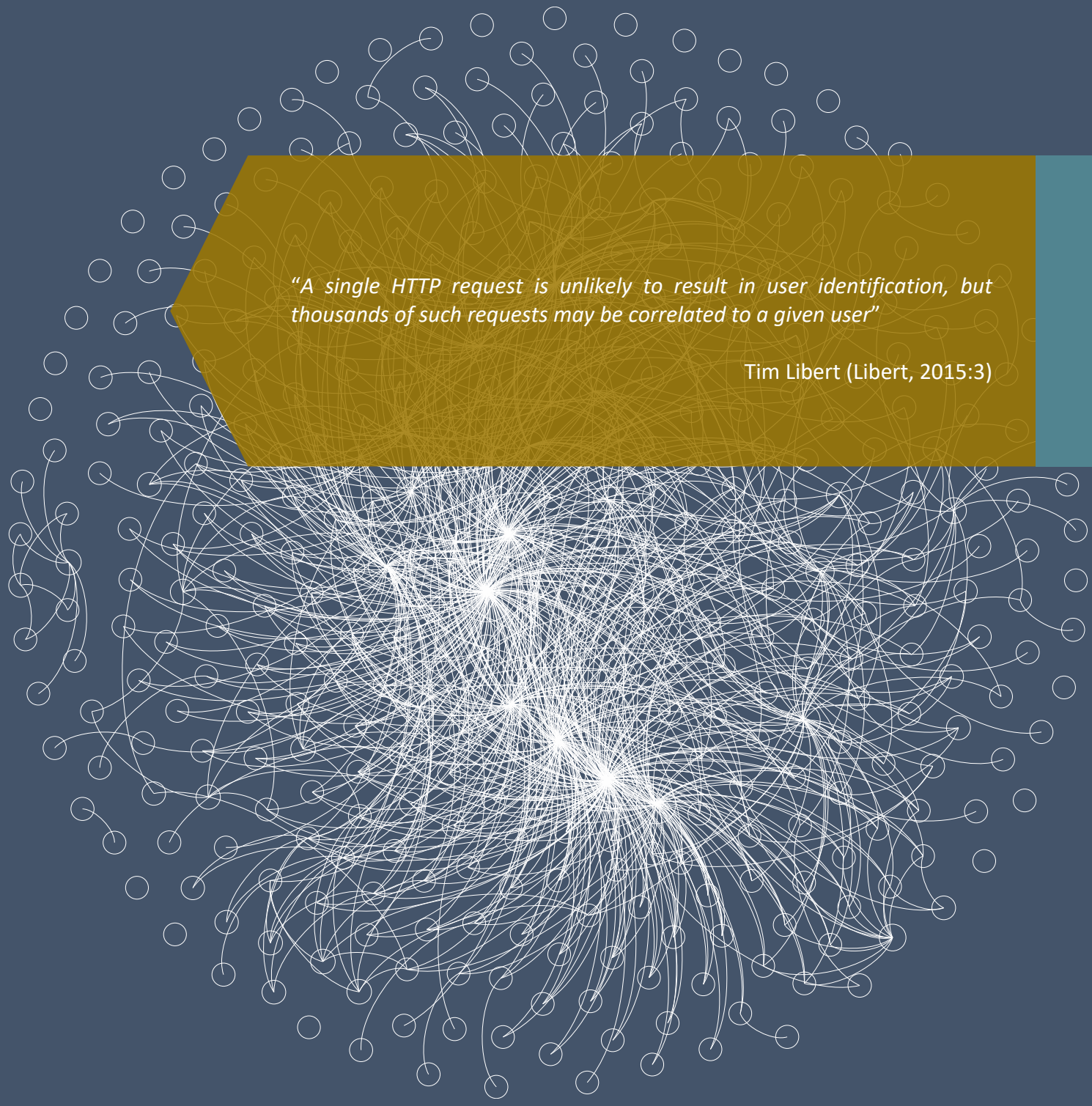
Danish public web pages through TPT might cause increased harm, as well as compromise the fundamental right to privacy, compared to the possible harm of the organizations who are being tracked (public institutions) and are tracking (trackers). Had our primary focus been a study of the different organizations that track, the harm balance might have been different, as this would be a more direct study of the organizations and the leakage of personal data would have been a secondary focus. This is however still up for discussion, as the balance of harm is subjective and contextual (Franzke, Bechmann, Zimmer, Ess and the Association of Internet Researchers, 2020).

The 2019 guidelines recently published should be understood and read in conjunction with the previous guidelines. It is therefore relevant to draw on all three versions. IRE 3.0 includes specific details about informed consent in data scraping studies. The concept of informed consent is not as relevant in our study, as we are not collecting information that is personally identifiable for human subjects. The data we are collecting from the Danish public web pages is publicly available, as the data consists of information about the TPSs on the web pages. However, it could be argued that WebXray is collecting data about the organizations who track, and the organizations should therefore be covered by some kind of consent (Franzke et al., 2020).

The overall discussion of ethics in our study is therefore a balance between publicly displaying the Danish public web pages using TPSs, hereby causing possible negative publicity. On the other hand, exposing the extent of TPT on Danish public web pages ensures awareness and possibly protection of human subjects from being subjected to data leakage from TPT.

This ethical discussion of balance and harm does therefore raise the overall question of regulatory focus within this arena. Do we need further awareness and regulation to protect the rights of human subjects concurrently with the increasing digital agenda within the Danish public sector?

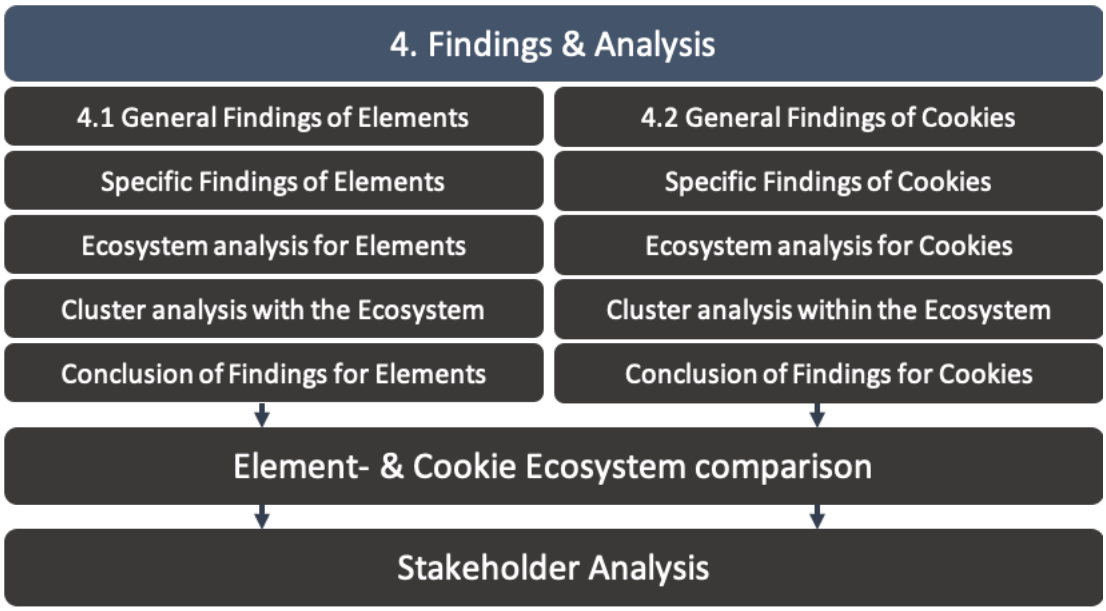
4. Findings & Analysis



"A single HTTP request is unlikely to result in user identification, but thousands of such requests may be correlated to a given user"

Tim Libert (Libert, 2015:3)

This section is an account of our findings, as well as an analysis of the results. Our findings and analysis section is structured according the figure below:



4.1 General Findings of Elements

The selected pages, which constitute the page list, were analyzed in March of 2020. During this analysis and data collection, WebXray encountered some errors, which meant that nine pages from the page list were not analyzed. These pages constitute 3% of the total number of pages on the page list, meaning that a total of 277 pages were analyzed.

In the analysis of the pages, WebXray managed to identify a total of 14147 requests for elements. Out of those requests, 2765 can be classified as third-party element requests, but since certain requests for the same third-party elements have been made on multiple pages, the number of unique third-party elements is 1393. In other words, requests for third-party elements constitute 20% of the total number of requests, whereas requests for domestic elements to own domains constitute 80%. Out of all of the 277 pages, 252 pages have requested elements from a TPS. This means that only 25 pages, or 9% of the analyzed pages, have not made third-party requests for elements, but have instead made requests to their own servers.

Of the top-level domains, or TLDs, "*.com*" is the most dominant as 75% of the third-party requests have been made to domains with this TLD, such as "googleapis.com". Other TLDs include "*.net*" (11%), "*.dk*" (6%), "*.io*" (6%) and "*.no*" (1%).

4.1.1 Third-Party Requests

The next section will seek to explain the requests for third-party elements on Danish public web pages, based on the data output from the WebXray analysis. The section focuses on the total set of 2765 third-party requests for elements, which have been registered.

Types of Elements - Element Characteristics

When it comes to determining the type of element that has been requested, the file extension of the element can be used. However, WebXray was unable to determine the file extension in 24% of the requests, which means that file extensions have been identified in 2107 third-party element requests. The table below illustrates the most common files extensions found in the third-party element requests.

Table 6: Top 10 most common file extensions in third-party requests

Rank	File extension	Count	Count %
1	js	918	44
2	woff2	274	13
3	css	227	11
4	jpg	219	10
5	aspx	143	7
6	png	110	5
7	html	73	3
8	gif	35	2
9	php	27	1
10	woff	13	1

As table 1 indicates, JavaScript (JS) is the most prevalent file extension and is found in 44% of the requests. Other common element types are the image elements, with the file extensions *JPG* (10%), *PNG* (5%) and *GIF* (2%), which are found in 17% of the element requests. The file extensions *WOFF* and *WOFF2*, which are used for text fonts on web pages, together constitute 14%, and *CSS*, which can be used to structure the page layout, fonts and colours, makes up 11%. Element types for displaying dynamic content on the page, *PHP* and *ASPX,* were also found and they constitute 8% together.

That JavaScript (JS) is the most common file extension is perhaps not that unusual, since JavaScript is a common language used for web page development. JavaScript elements can be used to make a web page interactive and responsive, such as modifying images or other elements based on the user interaction. JavaScript elements can have very different purposes, where some could be necessary for the essential function of the page and others could be on the page for non-essential reasons, such as for cosmetic or purely for tracking purposes (Libert, 2015).

This study does not look into the specifics of the requests for JavaScript elements, even though some JavaScript elements could potentially be a bigger threat to privacy than other types of elements (Libert, 2015). This is due to what Libert (2015) refers to as browser fingerprinting, where the JavaScript element contains code that, when loaded into the web page, registers the unique characteristics of the user's computer, which can then be used to identify the user in the future (Libert, 2015). Although this is a risk related to third-party JavaScript elements, the findings of this thesis do not provide any information on whether this could be the case on web pages of the Danish public sector, it is however an interesting finding.

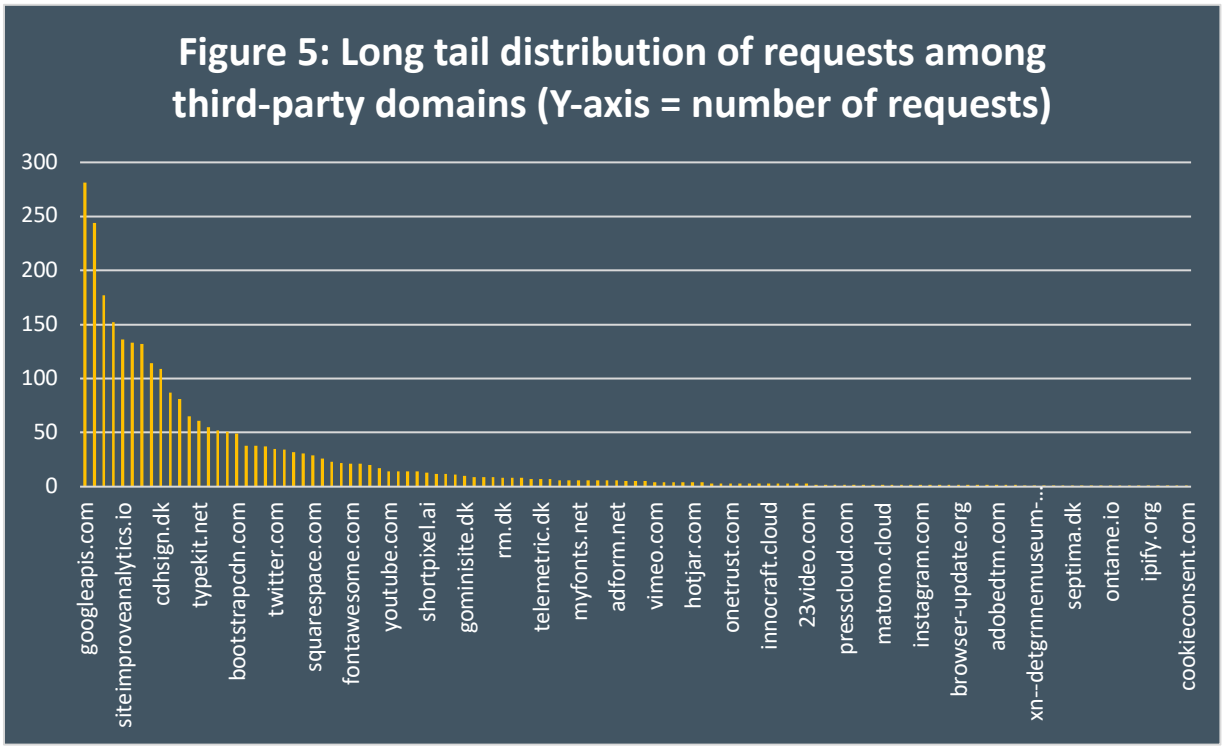
One example illustrating the complexity of this type of third-party element and that JavaScript elements can have functions other than making the page function correctly, is the element with the filename "trackingCode.js". This was found on the respective web pages for The Council of Appeal on Health and Safety at Work (Ankestyrelsen, Danish Ministry of Immigration, Integration and Housing (Udlændinge- og Integrationsministeriet) and The National Board of Social Services (Socialstyrelsen). The element comes from the domain "*fonts.net"*, which belongs to Monotype; an American company that provides web fonts (Fonts.com, 2020). From a domain titled "*fonts.ne*t", one could expect that the only elements to be requested would be fonts. However, given that this element has the *JS* file extension, i.e. JavaScript, it is not a font, or unlikely another element supporting the functionality of the font, as these mainly use the file extensions *WOFF*, *WOFF2 and CSS*.

Furthermore, the description in the filename *"trackingCode*" essentially implies that it has been loaded into the web page with the intention of tracking or doing something related to tracking. As aforementioned, this study does not intend to understand the specifics of JavaScript elements, but one could argue that it is suspicious that a TPS, which serves to provide fonts, is also loading in tracking-specific JavaScript elements.

4.1.2 Top Third-Party Domains

The 2765 third-party requests are shared between 117 different third-party domains. WebXray was able to determine the domains of all of the third-party element requests. The domains indicate where the element requests location. Our findings indicate that there is a long tail of third-party domains, where the most common and least common third-party domains are far apart in terms of requests, as Figure W shows.

Figure 5: Long tail distribution of requests among third-party domains



This figure is illustrated in the element ecosystem in section 4.1.5.

As figure 5 shows, there is an uneven distribution between the third-party domains and concentration of requests, as these are among the top domains. In order to better describe the third-party domains, the following section will focus on the top 15 third-party domains and subsequently on the bottom ranked third-party domains, which have only received a single or few requests. The domains vary in terms of the purpose that their elements serve. For example, some pages made requests to the domain "*cloudflare.com*", which is a TPS that provides internet security solutions. Meanwhile, some pages also made requests to domains associated with malware, which will be described in section 4.1.3.

The table below shows each third-party domain ranked according to how many element requests it has received.

Table 7: Characteristics of top third-party domains

Rank	Domain	Request count	Domain count %	Country
1	googleapis.com	281	10,2	US
2	gstatic.com	244	8,8	US
3	confirmit.com	177	6,4	NO
4	twimg.com	152	5,5	US
5	siteimproveanalytics.io	136	4,9	DK
6	google-analytics.com	133	4,8	US
7	cludo.com	132	4,8	DK
8	cookieinformation.com	114	4,1	DK
9	cdhsign.dk	109	3,9	DK
10	siteimproveanalytics.com	87	3,1	DK
11	cookiebot.com	81	2,9	DK
12	siteimprove.com	65	2,4	DK
13	typekit.net	61	2,2	US
14	fonts.net	55	2	US
15	cdninstagram.com	52	1,9	US

As seen from table 7, the domain "*googleapis.com*" (10,2%) is the third-party domain, which has received the most requests. This domain is American, as it is owned by Google, and it provides Google elements. The requested elements that were identified include fonts, Google Maps elements, Google Translate elements and elements related to the Google AJAX dynamic search box, which is an element that allows the page to have a search function, where the user can perform custom Google searches directly from the page.

Top Third-Party Domains for Analytical Elements

Some third-party domains mainly received requests for elements with analytical purposes. One such domain is "*google-analytics.com*". This domain is owned by Google and has likely been found, because the page, where the request was made from, uses Google Analytics, which is a program that allows the website owner to monitor and analyze user traffic (Google Analytics, 2020).

Requests to "*google-analytics.com*" were mainly made for JavaScript elements, however there were also some requests for elements with the GIF file extension, which indicates an image element. This is peculiar and somewhat unusual, since an analytical tool, such as Google Analytics, is not visible for the user when entering the web page. All the requested GIF elements from "*google-analytics.com*" has the file name *"_utm.gif"*. The same file name from the same third-party domain was found by Libert (2015). The image element is only 1x1 pixels in size, which is so small it is impossible for the user to see, and is used for tracking purposes only (Libert, 2015).

This example is somewhat comparable to the aforementioned finding of the "*trackingCode.js*" element from the "*Fonts.net*" domain, as neither of them seem to be central to the value that the respective TPSs are supposed to add to the web page, whether it would be page analytics or fonts. The element "_utm.gif" also shows the complexity of third-party elements, as it is disguised as an image element, when it is actually just used for tracking.

Another, and quite prominent TPS for analytical elements, is Siteimprove. Though the findings show three different third-party domains, namely "*siteimproveanalytics.io*" (4,9%), "*siteimproveanalytics.com*" (3,1%) and "*siteimprove.com*" (2,4%), they all belong to Siteimprove. Siteimprove is a Danish company that provides analytical software that allows the owner of the website to gain information about website performance, user activity and other related issues (Siteimprove, 2020a).

The reason that Siteimprove has three domains is not clear from the findings, but these domains differ largely on the types of requested elements. Requests made to "*siteimproveanalytics.io*" are solely for two elements with ASPX file extension, namely "*heat.aspx*" or "*image.aspx*", which indicates that it could be used for displaying dynamic content. Requests made to "*siteimproveanalytics.com*" are only for JavaScript elements, however there are various file names among these elements. Lastly, requests made to "*siteimprove.com*" are mainly for JavaScript elements and elements with the PNG file extension. The PNG elements are image elements, and these are used on web pages to display the icons for different types of cookies that the user can see, when he or she is selecting to allow or disallow cookies. Compared to the element "_utm.gif" from Google Analytics, image elements from the domains related to Siteimprove are actually visible to the user.

Top Third-Party Domains for Graphic Elements

Third-party domains that provide graphic elements are also found among the top third-party domains, including the third most used domain "*gstatic.com*" (8,8%). Like some other domains, "*gstatic.com*" is also owned by Google and the elements that were requested from here were mainly fonts and images, but also some JavaScript elements. Another third-party domain with requests for graphic elements is "*twimg.com*" (5,5%), which is an image hosting domain owned by Twitter. The third-party domains "*typekit.net*" (2,2%) and "*fonts.net*" (2%) are both providing fonts, however as aforementioned, JavaScript elements with tracking purposes were requested from "*fonts.net*". Furthermore, GIF elements that are 1x1 pixel in size, similar to the "*_utm.gif"* from Google Analytics, were requested from "*typekit.net*", which could indicate that this third-party domain is also providing elements solely for tracking purposes. The domain "*cdninstagram.com*" (1,9%) also falls under the category of providing graphic elements. Only JPG elements were requested from this domain, which is likely because this domain hosts images for the image sharing social media platform, Instagram, which is owned by Facebook. All the pages that contained elements from "*cdninstagram.com*", made at least five or more requests to this domain, which indicate that the images could be used to create an image collage or a kind of embedded Instagram feed.

Top Third-Party Domains for Elements with very Specific Purposes

Some of the top third-party domains provide elements for rather specific usage. The domain "*confirmit.com*" (8,8%), which is owned by the Norwegian company Confirmit, only received requests for elements that included "*digitalfeedback*" in the file name. Digital Feedback is a software from Confirmit that website owners can use to create user targeted pop-up surveys (Confirmit, 2020). Other examples of elements with very specific purposes, include the Danish based domain "*cludo.com*" (4,8%), which provides elements for site search optimization, and the domains "*cookiebot.com*" (2,9%) and "*cookieinformation.com*" (4,1%), which both provide elements that that create pop-up box for giving consent to cookies.

The domain "*cdhsign.dk*" (3,9%) also has a very specific purpose, as it provides elements that enable users with speech and hearing impairment to access the information on the page. Requests to this domain were made by multiple pages from Danish regions and municipalities. The domain appears to be a part of the digital tool, "Adgang Med Tegn" (Access with signs), which has been developed by the Danish Agency for Digitisation and other organizations related to speech and hearing impairment (Adgang Med Tegn, 2020). The domain "*cdhsign.dk*" received requests for JavaScript and CSS elements, which are used for dynamic content. This is consistent with how the tool is used,

as the user can select to replace certain words on the page with videos that show that particular word in sign language (Adgang med tegn, 2020).

Having such elements to help users with disabilities access information on a web page may also be an example of complying with strategies and regulation. In 2016, the European Union imposed a directive requiring member states to make web pages related to public services accessible for people with disabilities and the member states are to be evaluated in 2021 (European Commission, 2019).

One could argue that the elements from "*cdhsign.dk*" stand in contrast to certain elements, which only exist to the benefit of the TPS or web page owner, as elements from "*cdhsign.dk*" play a central role in enabling users with disabilities to access information on public web pages. Though the ethical aspects of the duality of elements will be considered later in the discussion section of this thesis, this example also shows how the most common element type in this study, JavaScript, can vary in both purpose and functionality.

4.1.3 Third-Party Domains with Few Requests Lacking Transparency

Whereas the prior section covered the third-party domains that received the most requests, this section will cover some interesting domains that received few requests. This is due to the fact that even though these domains receive few requests, they still obtain information about the user. As these domains have received few requests, it could indicate they are not related to commonly used TPSs by Danish public web pages, such as Google Analytics and Siteimprove. These TPSs might therefore track for more specific information. This segment will cover three third-party domains, which have only received requests on one or two pages and are relatively peculiar in nature.

The domain "*pulseadnetwork.com*" only received one request, which was for the element "*pix.html*" on the web page for the Billund Municipality. Through a WHOIS lookup, we found that the domain "*pulseadnetwork.com*" is registered under the American domain registrar GoDaddy, however, the owner of the domain was labelled as "*Domains by Proxy, LLC.*". Domains by Proxy is an American internet privacy service company and an affiliate of GoDaddy, that specializes in hiding information about the actual domain owner in WHOIS lookup results, which essentially reduces the quality of WHOIS searches (Caulfield, 2017). Subscribers to Domains by Proxy will be given proxy email addresses and phone numbers for the administrative-, technical- and billing contacts (Domains by Proxy, 2020), making them appear as owned by Domain by Proxy, and could therefore indicate someone trying to hide their true identity.

Domain by Proxy claims that the incentives for private registration include avoiding harassers and stalkers and shielding legitimate business endeavors (Domains by Proxy, 2020). However, Domains by Proxy has previously been involved in controversy, such as during the 2012 United States presidential election when owners of domains that antagonized Republican candidate, Rick Perry, such as "*buryperry.com*", were using Domains by Proxy to hide their identity (Rucker & Farnam, 2011).

The domain "*pulseadnetwork.com*" itself, has very little information available and trying to enter the domain in any browser does not yield any results. However, the domain appears on multiple IT related forums (Sharing Knowledge 2020; Cleanpcsolutions.com, 2016), where it is described as malware. However, the perhaps most interesting thing about this domain, is that we also found a cookie on the web page of the Billund Municipality from this domain, and like the element, the cookie was also only found on this page. That a cookie and element from the same domain are only found on the same web page, insinuates that there is a specific relation between the two, however our data output does not provide enough information to determine the characteristics of this relation. Even though the information about "*pulseadnetwork.com*" is scarce, it is arguably not ideal for a web page of a Danish municipality to request elements from domains that lack transparency to this extent and is categorized as malware of related it-forums.

The domain "*extreme-dm.com*" only received one request, which was for the element "*n3.g*" on the web page of the Randers Municipality. When entering the domain into an internet browser, it is redirected to "*extremetracking.com*". Though this website indicates that it is owned by a company called eXTReMe digital in the Netherlands, a WHOIS search for the domain did not yield any result and instead read "*Statutory Masking Enabled*". This is somewhat similar to the aforementioned example with Domains by Proxy, as it appears that the WHOIS search results about the domain owner have been intentionally hidden. "*Extremetracking.com*" provides tracking services for web page owners that can survey users visiting the page. The features include providing information about the specific user, including IP-address, internet service provider, browser, device operating system and screen resolution, as well as the user's path to the page and geographic location (Extreme Tracking, 2020).

Another uncommon third-party domain is "*moatads.com*". It received two requests for the element "*moatframe.js*" from the respective web pages for the Danish Home Guard (Hjemmeværnet) and for

Nota, which is an online library for people with reading disabilities. A WHOIS search for the domain does not yield any results and the result fields read "*Redacted for Privacy*", which is similar to the case with "*extreme-dm.org*". Though little information is available, prior studies have found that "*moatads.com*" is a domain that serves malware (Acker, Hausknecht & Sabelfeld, 2017; Nurse & Buckley, 2017). Given that "*moatads.com*" has been found to have a malicious purpose and considering the concerns about JavaScript elements' ability to perform browser fingerprinting (Libert (2015), it is concerning that this domain receives element requests from Danish public web pages.

4.1.4 Top Third-Party Elements

The prior section focused on the third-party requests, whereas this section seeks to describe the unique third-party elements. Whereas a third-party request can be made from different pages, requests may still be for the same element, such as the element "*iframe_api*" from YouTube, which has been requested on six different pages. The unique third-party elements have been ranked according to the number of pages that they have been requested on. The page count is based on the 277 analyzed pages. The table below describes the top unique third-party elements:

Table 8: Top unique third-party elements

Rank	File name	Domain	Company or program	Country	Page count	Page count %
1	css	googleapis.com	Google APIs	US	98	35,4
2	scenario	confirmit.com	Confirmit	NO	70	25,3
3	search-script.min.js	cludo.com	Cludo	DK	57	20,6
4	program	confirmit.com	Confirmit	NO	46	16,6
5	loader	confirmit.com	Confirmit	NO	46	16,6
6	collect	google-analytics.com	Google Analytics	US	41	14,8
7	analytics.js	google-analytics.com	Google Analytics	US	41	14,8
8	cookiesharingiframe.html	cookieinformation.com	Cookie Information A/S	DK	35	12,6
9	vt	googleapis.com	Google APIs	US	35	12,6
10	collect	google-analytics.com	Google Analytics	US	29	10,5

When looking at the file extensions for the top unique third-party elements, it became apparent that JavaScript is the dominant file extension, as it constitutes almost half of the 20 (only 10 is shown in table 8) most requested unique elements. However, it is also noticeable that some of the unique

third-party elements do not have a file extension included in the file name, such as the second ranked element "*scenario*" from Confirmit, which may explain why WebXray had difficulties with determining the file type of certain elements. Whereas Figure Z in segment 4.2.2 shows the total number of requests to a certain domain, Figure X can help us to illustrate the scope of the domains' different elements. Nine of the top 20 unique third-party elements have been made to domains owned by Google, while ten of them have been made to domains based in the United States. Furthermore, the fact that some of these elements have been requested by a large number of pages, indicate that they are a popular choice for many of the Danish public web pages, compared to elements from less common third-party domains, such as "*extreme-dm.org*", as mentioned in section 4.1.3. An example of a third-party domain with a popular element is "*cludo.com*", which hosts the third most requested unique element. "*Cludo.com*" is owned by the Danish company Cludo and provides search solutions for websites (Cludo, 2020).

Summary of the General Findings for Elements

The third-party elements found in this study have a variety of different file types and have been requested from various different domains. JavaScript is the most common file type, which could indicate that the majority of the requested elements are used for page functionality, rather than for design purposes, which images and fonts are typically used for. However, fully determining the purpose of an element based on the file extension is difficult and almost impossible, as we found elements with image file extensions, which turned out to be used solely for tracking purposes. Despite the varying purposes of JavaScript elements, we did find JavaScript elements, such as the element "*trackingCode.js*" from a domain that hosts fonts, which appeared to be used for tracking purposes. This essentially shows that it is difficult to determine the reason and purpose of certain elements, as they can perform many tasks on a web page or even be disguised as something that they are actually not, i.e. visible images.

In terms of the different third-party domains, we found that there is an uneven distribution of the requests, where these requests are concentrated among the top third-party domains. These top third-party domains are primarily owned by large corporations such as Twitter, Google and Siteimprove. While the top third-party domains are dominated by widely known corporations, the third-party domains that received few or single requests are owned by lesser known corporations and TPSs. Some of these third-party domains are registered through proxy domain registrars, which allows for hiding their actual corporate information in WHOIS searches, which indicate that they have an interest in remaining anonymous. Other third-party domains with a single or few requests appear

to be directly involved with malware, which is concerning. The analysis of the general findings for elements also covered the unique third-party elements, where the most requested unique elements come from domains used by programs for website analytics, such as Siteimprove and Google Analytics. Other highly requested elements appear to be used for displaying information about the respective web pages' use of cookies, as they were requested to domains associated with companies such as CookieBot.

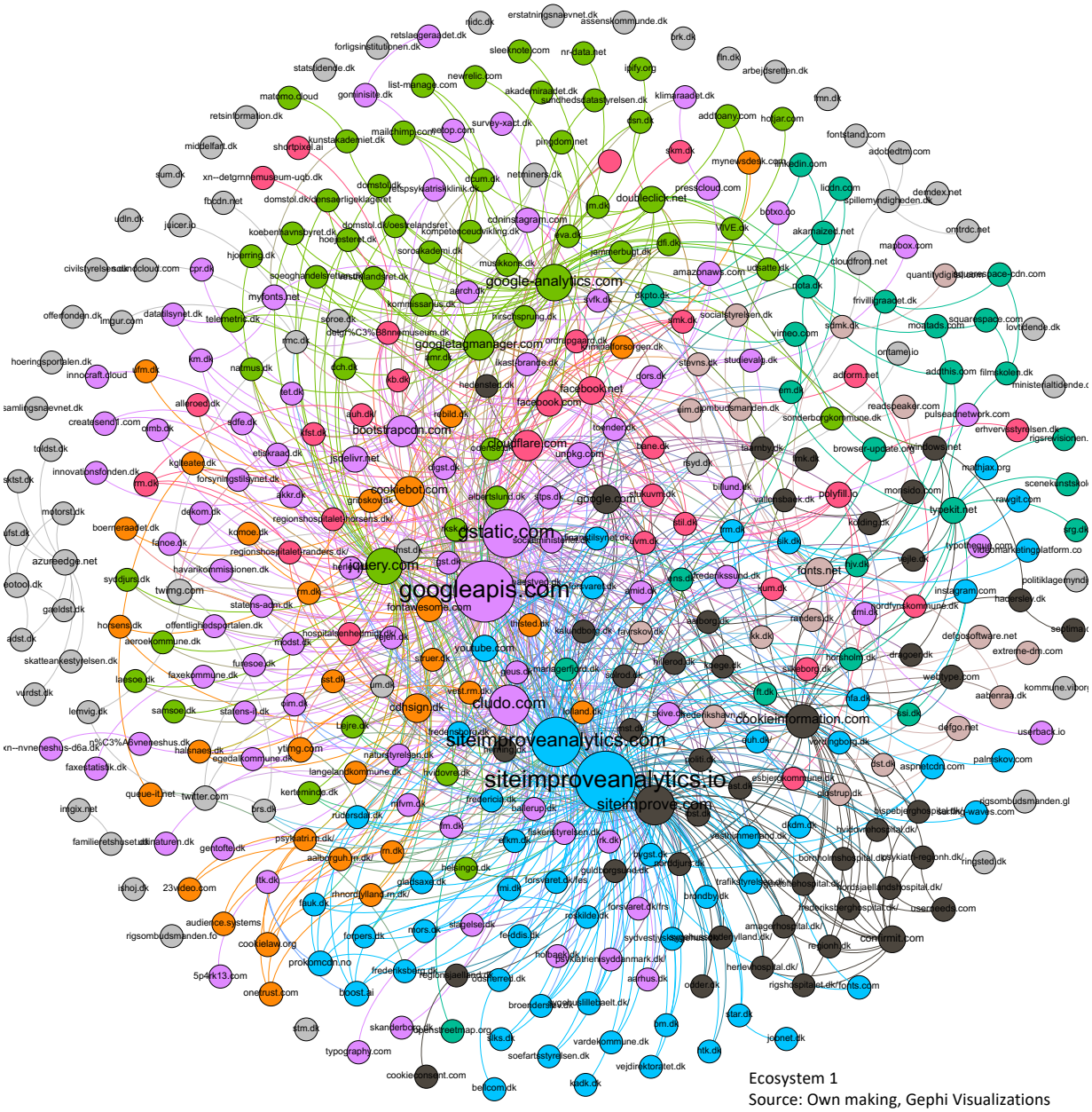
4.1.5 Ecosystem of Third-Party Element Domains

The prior sections of the findings and analysis have sought to describe the characteristics of the different third-party domains and elements found in this study. This section will focus on the relationship between the different third-party domains and web pages, through a mapping in an ecosystem of TPSs on Danish public web pages.

The ecosystem visualized below shows the connections between third-party domains and the different web pages they are connected to. A connection is when a web page makes a request for an element to a third-party domain. The sizes of the circles are proportionate in size compared to the number of connections a third-party domain has. The colours within the ecosystem are based on a modularity formula calculated by Gephi, as described in section 3 on Methodology. As visualized in the ecosystem, there are some really dominating third-party domains illustrated by the big circles. These are mainly domains from Siteimprove and Google, which are connected to an extensive number of public web pages. Both Siteimprove and Google have multiple domains, which are connected to many different pages in the ecosystem.

We are also seeing third-party domains that are connected to several pages, though not as extensive as the top third-party domains, and TPSs that are connected to only 1-3 pages. The small grey circles in the ecosystem are pages that did not make element requests to any third-party domains. We looked into a number of these pages, in order to understand why they do not make requests to third-party domains and found no evidence or pattern that could explain this.

The ecosystem below illustrates the noticeable complexity with many different third-party domains varying heavily in size and number of connections. This complexity indicates a rather unstructured use of TPSs.

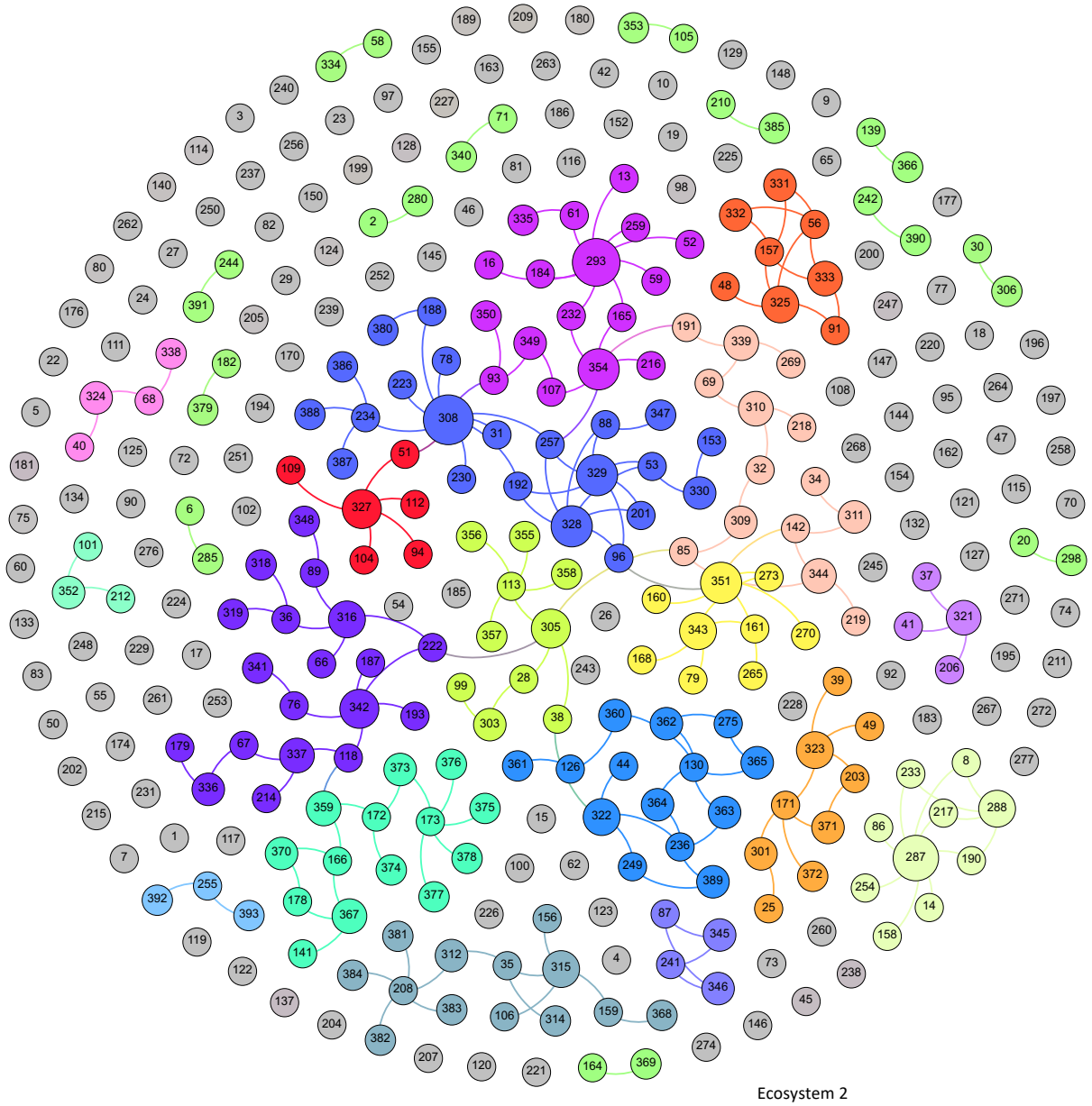


4.1.6 Clusters within the Ecosystem

The ecosystem visualized below is based on the same criteria as the ecosystem above. The difference between the two ecosystems is that we removed the top trackers, i.e. third-party domains with 10+ connections, to get a better understanding of the underlying patterns within the ecosystem. This allowed us, through the modularity formula, to create clusters within the ecosystem, which we visualized with the different colours.

The modularity of colours shows a clear division of clusters. All the small grey circles are Danish pages that do not make element requests to third-party domains, the pages that only make requests

to a single third-party domain are coloured green, etc. The modularity colour therefore shows the division of clusters within the ecosystems. The ecosystem does not contain names, but rather IDs representing the names, in order to better show the connecting lines between the circles. The clusters of relevance will be analyzed individually below, where the third-party domains and page-names will be listed together with their respective IDs.



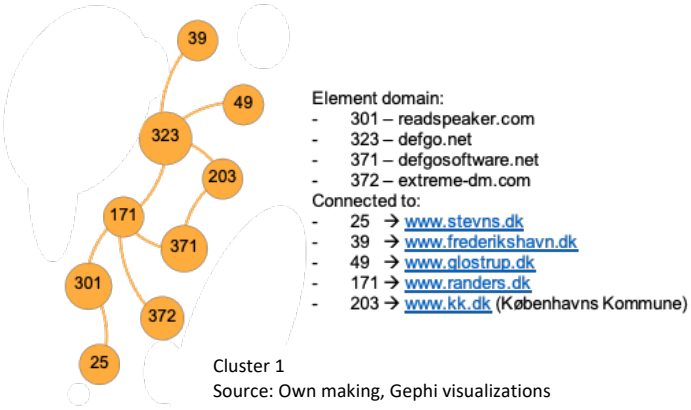
Ecosystem 2
Source: Own making; Gephi Visualizations

Clusters on Municipal, Regional and National Levels

The following section will cover the clusters that have been formed based on pages that represent public entities on municipal-, regional- and national levels respectively.

Municipal Level

Despite the high level of arbitrariness, some clusters consist of somewhat related pages. The following cluster is based on requests to multiple domains, while "*defgo.net*" is the most interesting one. Pages of four municipalities are requesting this domain, however, these municipalities are not similar in terms of geography or size. A visualization of the cluster and legend for the identification numbers can be seen below.



Cluster 1
Source: Own making, Gephi visualizations

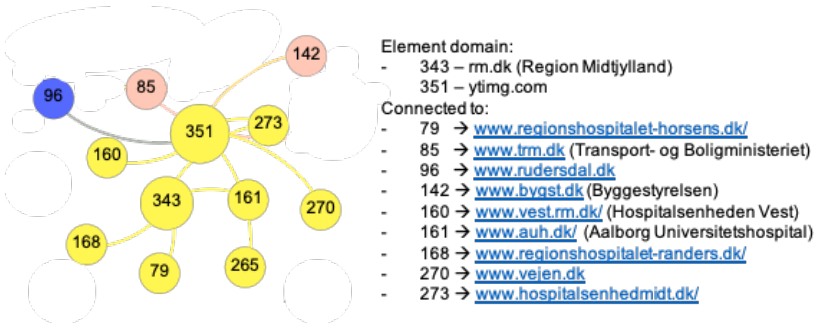
The domain "*defgo.net*" is owned by the Danish company Defgo, which offers solutions to present surveys to page visitors (Defgo, 2020). As only four pages requested elements from this domain, it could indicate that this solution could be especially fitting for municipalities, but given the cluster's size, it is not conclusive. The domain "*defgo.net*" also set cookies on the pages for the municipalities of Copenhagen and Randers, but not on the remaining two pages. The output from WebXray does not provide a probable explanation to this.

Another cluster consisting of pages representing municipalities, is the cluster based on the requests to multiple domains, however most importantly "*boost.ai*", which is owned by the Norwegian company Boost.ai, and "*prokomcdn.com*", owned by a Prokom, which as a partner of Boost.ai. Requests to these domains have only been requested on pages representing municipalities in Eastern Denmark. A cookie cluster consisting of the same pages and the "*boost.ai*" domain has also been found, which is why the cluster will be visualized and analyzed further in section 4.2.3.

Regional Level

Whereas the previous clusters consisted of pages on a municipal level, we also encountered a cluster, which contained only pages on a regional level. This cluster is based on element requests to two domains. The first domain is "*ytimg.com*" and is a third-party domain from YouTube for hosting

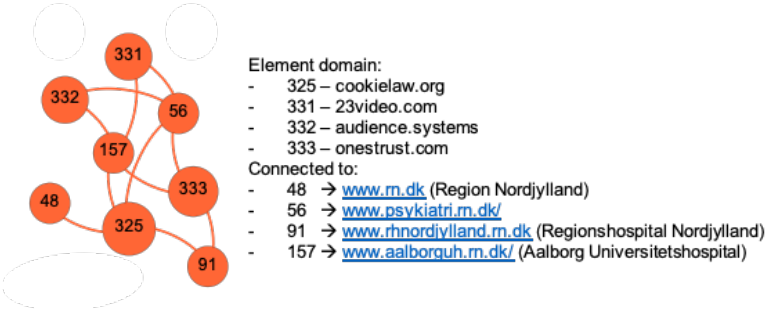
thumbnail images. The second, and more interesting domain in this case is "*rm.dk*", which is also the domain for the Central Denmark Region (Region Midtjylland), and the pages in this cluster are all referring to the four hospitals that are administered by this region. Please note that circle 343 (rm.dk) is not connected to circle 351 (ytimg.com), but rather circle 273 (hospitalsenhedmidt.dk).



Cluster 2
Source: Own making, Gephi visualizations

The same unique element was requested across the pages and it has the PNG file extension, which indicates that it is an image. This cluster is an example of a public entity, i.e. Central Denmark Region, that appears as a TPS. However, the pages are all representing entities, which fall under the administration of this region, making it unlike other private TPSs found in this study, because there is a clear organizational relationship in this cluster.

Pages for entities, which are all administered by North Jutland Region (Region Nordjylland), also formed a cluster, as these pages more or less only requested elements from the same three third-party domains.

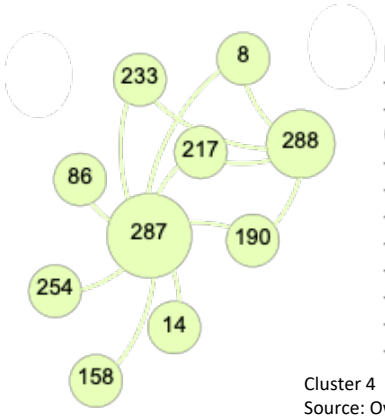


Cluster 3
Source: Own making, Gephi visualizations

This could indicate that the region provides a set of guidelines of which TPSs to use. The reason for the consistent usage of TPSs, could be that the pages have different subdomains and identical domains, such as "*psykiatri.rn.dk*" and "*aalborguh.rn.dk*". However, regardless of this, it still ensures a more controlled and limited use of different TPSs, which could ultimately be to the benefit of the user.

National Level

An example of a cluster, where the pages could be related on a national level, is the cluster based on element requests to the domain "*videotool.dk*".



Element domain:
- 287 – axuredge.net
- 288 – videotool.dk

Connected to:
- 8 → www.ufst.dk (Udviklings- og forenklingsstyrelsen)
- 14 → www.vurdst.dk (Vurderingsstyrelsen)
- 86 → www.toldst.dk (Toldstyrelsen)
- 158 → www.skst.dk (Skattestyrelsen)
- 190 → www.motorst.dk (Motorstyrelsen)
- 217 → www.gaeldst.dk (Gældsstyrelsen)
- 233 → www.adst.dk (Administrations- og Servicestyrelsen)
- 254 → www.skatteankestyrelsen.dk

Cluster 4
Source: Own making, Gephi visualizations

Contrary to the majority of the clusters in this ecosystem, the pages in this cluster are very closely related as they are all pages of government agencies under the Ministry of Taxation. Furthermore, these pages are the same as in the cluster based on cookies from the domain "*skat.dk*", which will be described in section 4.2.4.

The domain "*videotool.dk*" is owned by VideoTool, which is a Danish company that provides solutions for hosting and publishing videos. VideoTool claims to have other entities of the public administration as clients, such as Aarhus Municipality and the Ministry of Education (VideoTool, 2020), however WebXray did not register any requests to the domain "*videotool.dk*" from the pages affiliated with these entities.

Furthermore, VideoTool claims to be GDPR-compliant and is a member of the SKI program, which means that the company is approved to sell solutions to entities in the public sector and thus engage in private-public partnerships. This indicates that the entities behind these web pages have considered the benefits of using such a TPS, compared to other TPSs that also provide video solutions, such as YouTube or Vimeo. The other third-party domain that only received requests from the pages within this cluster is "*azuredge.net*", which is a service from Microsoft that allows web page owners to securely host content (Microsoft, 2020). While these factors indicate that the Ministry of Taxation has deliberately chosen a secure TPS for certain pages, it also raises the question as to why this is not the case for the rest of the pages related to the entities under the Ministry of Taxation

or even the total public sector. This arbitrary and inconsistent use of TPSs will be discussed further in section 5, which covers the theoretical implications and discussions.

While VideoTool appears in a cluster, where we found a relation between the pages, we also found clusters based on third-party video content, where the pages have little to none relation. For example, one cluster containing pages that sent requests to the domain "*vimeo.com*", is referring to Vimeo, which is an online video sharing platform. The requests to this domain were all made on pages for entities on the national level, such as the Danish Energy Agency and the Ministry for Industry, Business and Financial Affairs. Apart from the page of one agency that lies under the Ministry for Industry, Business and Financial Affairs, there does not seem to be any relation between the pages. However, when comparing this cluster to the cluster based on VideoTool, it becomes apparent that different ministries and agencies are using different TPSs for the purpose of displaying video content. This enforces our argument that there is an arbitrary use of TPSs and that the use of these is decided by the different entities that operate their own web pages.

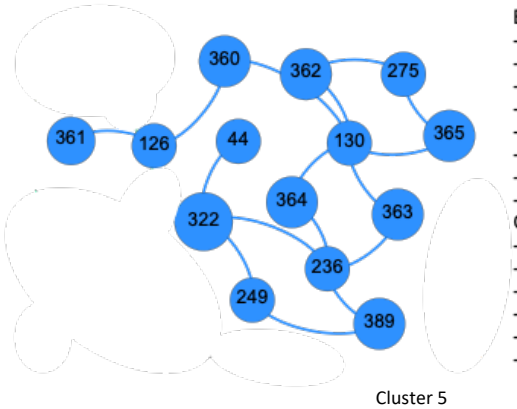
Many Third-Party Services for the Same Purposes

As it can be seen by looking at the VideoTool and Vimeo clusters, the pages analyzed in this study are occasionally using different TPSs for the same purposes, such as for displaying video content. However, displaying video content is not the only purpose where different TPSs have been used. Other clusters were formed on the basis of pages using TPSs for fonts. The third-party domains "*fonts.com*", "*webtype.com*", "*myfonts.ne*t", "*typhoteque.com*" and "*fontawesome.com*" are all owned by TPSs that provide fonts and they all form their own clusters. Furthermore, there is little to no relation between the pages within each of these clusters, which makes it difficult to provide an explanation as to why these different TPSs have been used to provide fonts.

The cluster based on element requests made to the domain "*webtype.com*", consists of three pages, which are all representing municipalities, which could indicate some relation, though this is far from conclusive. The pages in the cluster based on the domain "*myfonts.net*" are relatively unrelated, as this cluster consists of the pages of the Danish Data Protection Agency and the Royal Danish Academy of Fine Arts, as well as two municipalities. Especially the pages in the latter cluster do not appear to have similarities, which enforces the argument that the usage of TPSs in this study is arbitrary.

Excessive Use of Third-Party Services

Some clusters were formed in an almost 'reverse' fashion, as these are not clusters where there is a single TPS present on multiple pages, but rather clusters based on a single page that send requests to multiple third-party domains. One such cluster includes the page "*nota.dk*", which is the page for a public library for people with reading disabilities.



Element domain:
- 322 – vimeo.com
- 360 – browser-update.org
- 361 – openstreetmap.org
- 362 – addthis.com
- 363 – licdn.com
- 364 – linkedin.com
- 365 – moatads.com
- 389 – akamaized.net

Connected to:
- 44 → www.ens.dk (Energistyrelsen)
- 126 → www.mariagerfjord.dk
- 130 → www.nota.dk (Online bibliotek)
- 236 → www.dkpto.dk (Patent- og Varemærkestyrlesen)
- 249 → www.em.dk (Erhvervsministeriet)
- 275 → www.hjv.dk (Hjemmeværnet)

Cluster 5
Source: Own making, Gephi Visualizations

Given that the top third-party domains have been removed in this ecosystem, this cluster shows an excessive use of minor TPSs. The third-party domains in this cluster deliver elements to fulfill various tasks. These include sharing buttons for social media, LinkedIn plugins and a special plugin that informs the page visitor to perform a browser update. The purpose of the third-party domain "*moatads.com*" is unknown, but it is likely associated with malware, as mentioned in section 4.1.3.

Another page with notable usage of lesser common TPSs is "*sik.dk*", which is the page for the Danish Safety Technology Authority (Sikkerhedsstyrelsen). This page makes requests to third-party domains, which have very different purposes. One of the domains is used for displaying video, as the domain "*videomarketingplatform.co*" is owned by the video hosting service, 23video.

Though this is a less common TPS for this purpose and other pages have used similar services such as Vimeo and VideoTool. Another domain, namely "*mathjax.org*", received requests for JavaScript elements, which help display mathematical equations on the page (Mathjax, 2020). Why this is necessary or relevant on the page for the Danish Safety Technology Authority is unknown. Requests were also made to "*rawgit.com*", which is unusual since the TPS behind this domain, RawGit, suspended operations in 2018, because RawGit had developed into a vehicle for distributing malware by its users (RawGit, 2018).

These two pages show not only an excessive use of TPSs, but an excessive use of TPSs, which are not commonly used throughout the ecosystem. Like the clusters with pages that have little to none relation with each other, these clusters show that there is perhaps a lack of strategy or guidelines regarding the usage of TPSs, when requests are made to less common third-party domains, which in certain cases are related to defunct or malicious TPSs.

4.1.7 Element Ecosystem Summary

The element ecosystem is primarily characterized by a large number of the clusters exhibiting an arbitrary use of TPSs. These clusters consist of page domains from various entities such as municipalities, agencies and ministries, which are rarely related. The clusters in this ecosystem are a strong indication that the implementation and usage of third-party elements is controlled by the individual entities that administer their own pages. This type of decentralization can also be seen when looking at the purpose of the domains. For example, in the context of graphical elements, such as videos, images and fonts, the pages are using various different domains for this purpose.

Though there is a high level of arbitrariness, some clusters consist of pages that are related. This is the case in clusters with private third-party domains, but also for public third-party domains. Some clusters were formed in almost 'reverse' fashion, as these are based on a single page that requests elements from many different domains, which receive zero or few other requests. The third-party domains usually have elements for very specific purposes and some of these third-party domains are even likely malware.

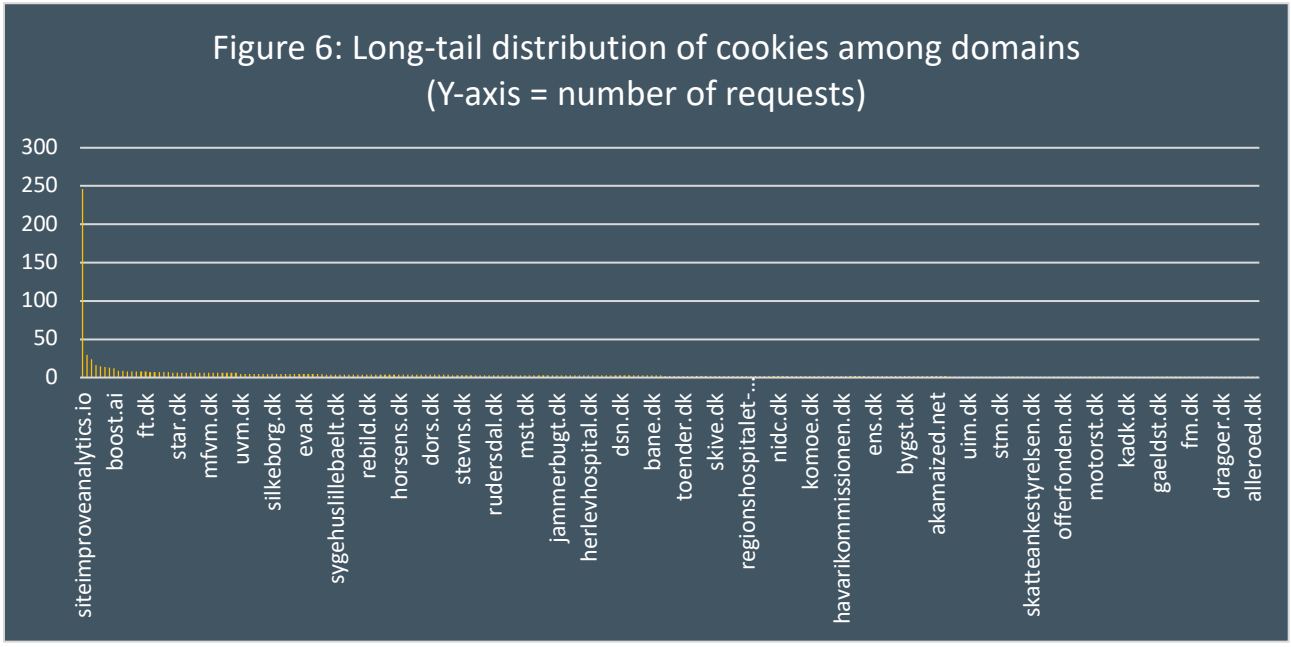
4.2 General Findings of Cookies

WebXray recorded 1096 instances of cookies. However, these cookies were only set on certain pages, as WebXray did not register cookies on 36 pages, which constitutes 13% of the 277 analyzed web pages. Some of the registered cookies came from third-party domains, however some cookies came from the same domain as the respective web page that the cookie(s) were found on, which emphasizes the complexity of the ecosystem. Furthermore, certain domains that appear on the page list, such "r*egionh.dk*", set cookies on other pages than its own, i.e. "https://www.regionh.dk/". These types of domains set cookies on pages that can be clustered based on their theme, which will be covered in section 4.2.3.

4.2.1 Cookie Domains

This section describes the domains that set cookies. The domains have been ranked according to the number of times the domain has set a cookie. There are 261 different domains that have set cookies, however some of these cookies may be of the same variant. One example of this is the cookie "*_ga*", which is a cookie that originates from Google Analytics, despite that the cookie-setting domain usually appears to be the same as the domain of the page, which it was found on. An example of this is on the page "https://www.*odense.d*k", where the "*_ga*" cookie is set by the domain "*odense.dk*". This also partially explains the long-tail of cookie-setting domains, which is illustrated in the chart below.

Figure 6: Long-tail distribution of cookies among domains



Source: Own making, Excel

This figure is illustrated in the cookie ecosystem in section 4.2.3.

In order to get a better insight into these domains, the next section will first highlight the ten most common domains and secondly selected domains that only set cookies on a small number of pages, where the cookie domain is different from the domain of the web page.

The table below shows the ten most common domains in ranked order.

Table 9: Top cookie domains

Rank	Domain	Count	Count %	Country
1	siteimproveanalytics.io	246	22,4	DK
2	domstol.dk	30	2,7	DK
3	rn.dk	24	2,2	DK
4	rm.dk	16	1,5	DK
5	youtube.com	15	1,4	US
6	linkedin.com	14	1,3	US
7	regionh.dk	13	1,2	DK
8	boost.ai	12	1,1	NO
9	frivilligraadet.dk	9	0,8	DK
10	facebook.com	9	0,8	DK

"*Siteimproveanalytics.io*" (22,4%) from Siteimprove is by far the most common cookie-setting domain. Siteimprove also owns the domain "*siteimprove.com*" (0,2%). That Siteimprove uses multiple domains is also the case of element-trackers, where it uses three different domains to provide elements, as explained in section 4.1.2. The domain "*siteimprove.com*" sets the same types of cookies as "*siteimproveanalytics.io*", which means that there is nothing in the results that points to an explanation for the fact that Siteimprove sets cookies from both domains.

The cookies from the domains "*linkedin.com*" (1,3%), from LinkedIn, and "*facebook.com*" (0,8%), from Facebook, are examples of cookies from social media platforms. The cookies from LinkedIn found in this study, are special third-party cookies, which are only on pages where LinkedIn is used as a TPS (LinkedIn, 2019). According to LinkedIn's cookie policy as of January 2020, these LinkedIn cookies can be set in the user's browser when the user visits a web page with a LinkedIn Plugin, such as a LinkedIn shortcut button (LinkedIn, 2020). Both of the pages, where cookies from the domain "*linkedin.com*" were found, have buttons with a LinkedIn logo that takes the user directly to the LinkedIn website, which likely explains the cookies' presence on these pages. The cookies from Facebook also follow the same pattern, as Facebook cookies can also be found on web pages that contain a Facebook plugin (Facebook, 2020). Essentially, the value that plugins for social media platforms create for an organization's web page, is the ability for the user to access the organization's social media profile faster, than through alternative ways. However, the benefit comes at the cost of

user privacy, as third-party cookies from these social media platforms are set in the browser of the user, without the user being necessarily aware of it. This dilemma of improved services vs. limited privacy will be discussed further in section 5.

Google is also represented among the most cookie-setting domains, as "*youtube.com*" (1,4%) ranks fifth. There are three different cookies set by this domain across five pages. The purposes of these cookies include enabling tracking of the user based on geographic location, estimate the user's bandwidth on pages with embedded YouTube videos and to register which YouTube videos that the user has watched (Cookiebot, 2020a).

The domain "*boost.ai*" is the only domain among the most cookie-setting domains, which is not related to a company in Denmark or the United States of America. This domain is owned by the Norwegian company Boost.ai, which is specialized in creating virtual agent and chat solutions for websites in the public sector, in order to better respond to inquiries and questions from citizens (Boost.ai, 2020).

Domains that Set Few Cookies

The prior section described the most common cookie-setting domains. This section will focus on the domains with only cookies on a single web page, which it does not share an identical domain name with.

The domain "*demdex.net*" set the cookie "*demdex*" on the web page for the Danish Gambling Authority (Spillemyndigheden). An element with an unidentified file extension from this domain was also found on this page, presumably in relation to the cookie. The domain "*demdex.net*" is owned by Adobe and is used to support Adobe Audience Manager, which is a tool for website owners to collect commercially relevant information about visitors with the purpose of serving targeted advertising (Adobe, 2020). The cookie "*demdex*" assigns a unique ID to the user and in that way helps Adobe Audience Manager perform its tasks. Another rare cookie-setting domain related to analytical tools, is "*nr-data.net*" setting the cookie "*JSESSIONID*" on the web page of the Danish Evaluation Institute (Danmarks Evalueringsinstitut). This domain is owned by New Relic, which is an American company that provides website analytical software for website owners. Cookies from the domain "*nr-data.net*" are used to transfer information about the user from the page that the user visits to New Relic's data collection servers (New Relic, 2020). Essentially, New Relic's cookies are relatively similar to those of Siteimprove and Google Analytics, since they enable the website owner's analytical programs to

function correctly. Another cookie which has also only been set on one page and supports an analytical program, is the "*sp*" cookie from Ontame.io. Ontame.io is a Danish company that provides software for analyzing recruitment campaigns (Ontame.io, 2020).

A rare third-party domain that has set a cookie, which is used for supporting embedded third-party media, is the domain "*soundcloud.com*". This domain set the cookie "*sc_anonymous_id*" on the web page of the Rythmic Music Conservatory (Rytmisk Musikkonservatorium). The domain is owned by the American music streaming service SoundCloud and the cookie is used to identify a user that visits a web page that has a SoundCloud music player embedded (SoundCloud, 2018). The web page where the cookie was set, also requested the element "*player.js*" from SoundCloud, which indicates that there is a media player plugin of some sort on the page.

The domain "*jobnet.dk*" set a cookie on the web page for the Danish Agency for Labour Market and Recruitment, *or STAR* (Styrelsen for Arbejdsmarked og Rekruttering). The domain is owned by Jobnet, which is an online recruitment portal, where citizens in Denmark can apply for vacant positions (Jobnet, 2020) and it is operated by STAR and Local Government Denmark (Kommunernes Landsforening). The web page for STAR also sent two requests to "*jobnet.com*" for elements labelled as "*analytics*", indicating analytical purposes are the explanation for Jobnet's cookie on the web page of STAR.

The domain "*pulseadnetwork.com*" has previously been described in section 4.1.3, as a third-party domain related to malware. This domain set the cookie "*accompat*" on the web page of Billund Municipality, which is the same page that made an element request to "*pulseadnetwork.com*". As little to none information about the purpose of this cookie is available, it seems rather peculiar that this cookie is set on the web page of the Billund Municipality.

4.2.2 Top Cookies

The table below shows the ten most common cookies found in this study. The cookies are ranked according to how many times they have been found in the WebXray analysis. As no cookies have been found multiple times on the same page, ranking them according to the number of pages that they were found on, would yield the same result. Out of the total 1096 findings of cookies, there are 253 different cookies across the pages.

Table 10: Most requested cookies

Rank	Name	Origin	Count	Count %
1	siteimproveses	Siteimprove	125	11,4
2	AWSELB	Siteimprove/Amazon Web Services	125	11,4
3	nmstat	Siteimprove	121	11
4	ASP.NET_SessionId	Siteimprove	70	6,4
5	_gid	Google Analytics	42	3,8
6	_ga	Google Analytics	42	3,8
7	_gat	Google Analytics	26	2,4
8	has_js	Drupal	23	2,1
9	szcib	Unknown	18	1,6
10	ARRAffinity	Siteimprove	15	1,4

Google is strongly represented among the top ten most common cookies. As mentioned in section 4.2.1, the cookie "*_ga*" is from Google Analytics, but is usually set by the domain of the web page that it is found on. Google Analytics uses this cookie to identify users by giving them a unique ID, which can then for example be used to determine if it is a recurring user (Cookiepedia, 2020). This cookie is also found together with the cookies "*_gat*" and "*_gid*", which are also cookies that are necessary for Google analytics to function correctly. Certain information about these cookies, including the domain, is customizable for the web page owner, which explains why the domains for these cookies vary.

In all cases, but one, the domain for these cookies from Google Analytics were identical to the domains of the web page the cookies were found on. However, on one of the analyzed web pages, the domain of these three cookies did not match the domain of the web page. On the web page for the Odense University Hospital, that has the page domain "*ouh.dk*", the cookies "*_ga*", "*_gat*" and "*_gid*" refer to the domain *"surfing-waves.com*".

From the web page "*https://surfing-waves.com/*", Surfing-Waves appears to be an online forum and news portal dedicated to sport surfing, where users can gain access to surf maps, chat rooms, surf shops, surf guides and other surfing related topics (Surfing Waves, 2020a). A WHOIS search for the domain "*surfing-waves.com*" did not yield any results regarding the ownership of the domain. However, it showed that the domain registrant is a company called WhoisGuard and it is located in

Panama. WhoisGuard is an internet privacy service, similar to Domains by Proxy (See section 4.1.3) capable of hiding information about the actual domain owner in WHOIS searches, by providing proxy information (WhoisGuard, 2020).

There are also two elements from the domain "*surfing-waves.com*" on the web page "*ouh.dk*". These two are dynamic content elements, with the *JS* and *PHP* file extensions, and appear to be used to embed a free news feed widget from Surfing Waves (Surfing Waves, 2020b) into the web page. Though Surfing Waves receives requests for its own elements from "*ouh.dk*", our results do not provide any immediate explanation as to why cookies related to the functionality of Google Analytics is referring to the third-party domain "surfing-waves.com".

Siteimprove also has a strong cookie presence, as we found that the company is behind the five most common cookies. These are used to collect information about the visiting user. This information is then used by Siteimprove's analytics software to analyze website performance and user behavior (Siteimprove, 2020b). Though the cookie "*AWSELB*" refers back to Siteimprove, it actually originates from Amazon Web Services. This is due to the fact that Siteimprove uses Amazon Web Services in its own solutions, however the results do not indicate whether user data is also leaked to Amazon Web Services.

A common cookie, which is not directly related to any analytical tool, is the "*has_js*" cookie from Drupal. The purpose of this cookie is implied in the name, as it checks if the user has JavaScript enabled. This is quite relevant, since the majority of the third-party elements found in this study are JavaScript elements (See section 4.1.1) and thus it requires the user to enable JavaScripts for this file type for them to work.

Looking at the ten most common cookies, most of them are on the web pages for analytical purposes, where they support analytical tools for the web page owners, such as Google Analytics and Siteimprove. That the cookies mostly originate from Google and Siteimprove, indicates that these providers of website analytics tools are popular choices of web pages of the Danish public administration.

Cookie General Findings Summary

The majority of cookies found in this study can be related to the use of third-party analytical programs by the web page owners, which are the different administrative entities in the Danish public sector.

This could be an indication that the tracking abilities possessed by cookies are mainly used to optimize performances on these web pages. Some of these cookies that are related to third-party analytical programs, such as the cookies for Google Analytics, refer to the domain of the respective web pages and not to a third-party domain, which is also an indication for the complexity of these tracking mechanisms.

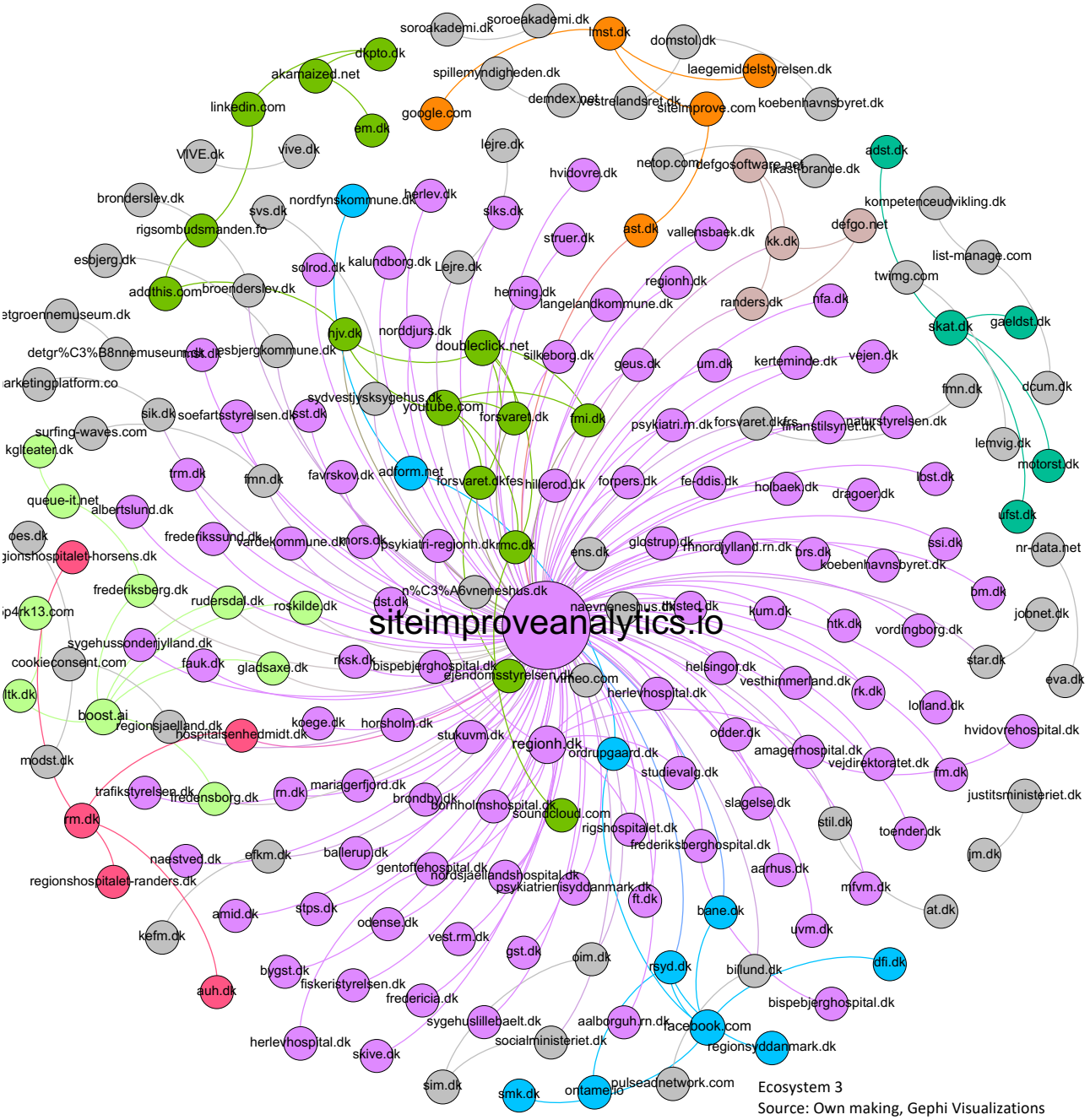
We also found cookies that refer back to third-party domains. Some of these were set on the pages as a result of embedded content, such as share buttons from Facebook and music players from SoundCloud. In one case, a cookie from a domain related to malware was also found. The fact that third-party cookies are set and activated instantly, when a user enters the web page, would actually mean that the usage of cookies from TPSs on some of the analyzed pages, is not in compliance with the guidelines for third-party cookies from the Danish Business Authority (Erhvervsstyrelsen, 2017), as mentioned in section 1.2. Another interesting fact is that some page domains set cookies on other pages than their own. These will be covered in the following section.

4.2.3 Ecosystem of Cookies

The ecosystem visualized below illustrates the connections between domains that set cookies, which can come from either private third-party domains such as "*siteimprovenanalytics.io*", or public page domains, such as "*regionh.dk*". We did not actively accept any cookies, when we analyzed the pages using WebXray, meaning that these cookies were set and activated without consent. The size and colour of the circles are based on the same factors described in the element ecosystem.

The ecosystem is dominated by one TPS, Siteimprove, and its domain "s*iteimproveanalytics.io"*. The size of the other circles is distorted by the size of this domain, as it is such an extensive TPS and the proportionate size of the other TPSs is therefore low. Apart from Siteimprove, there are also TPSs that set cookies on several pages and some only set cookies on 1-3 different pages.

The ecosystem for cookies is considerably complex, but not as complex as the ecosystem for elements. However, it is interesting that cookies are set on an extensive part of the pages, taken into consideration that we did not accept any of the cookies and the fact that all the pages are operated by the public administration. Just as the element ecosystem, the cookie ecosystem contains everything from TPSs with extensive tracking to TPSs that only track on a single or few pages.

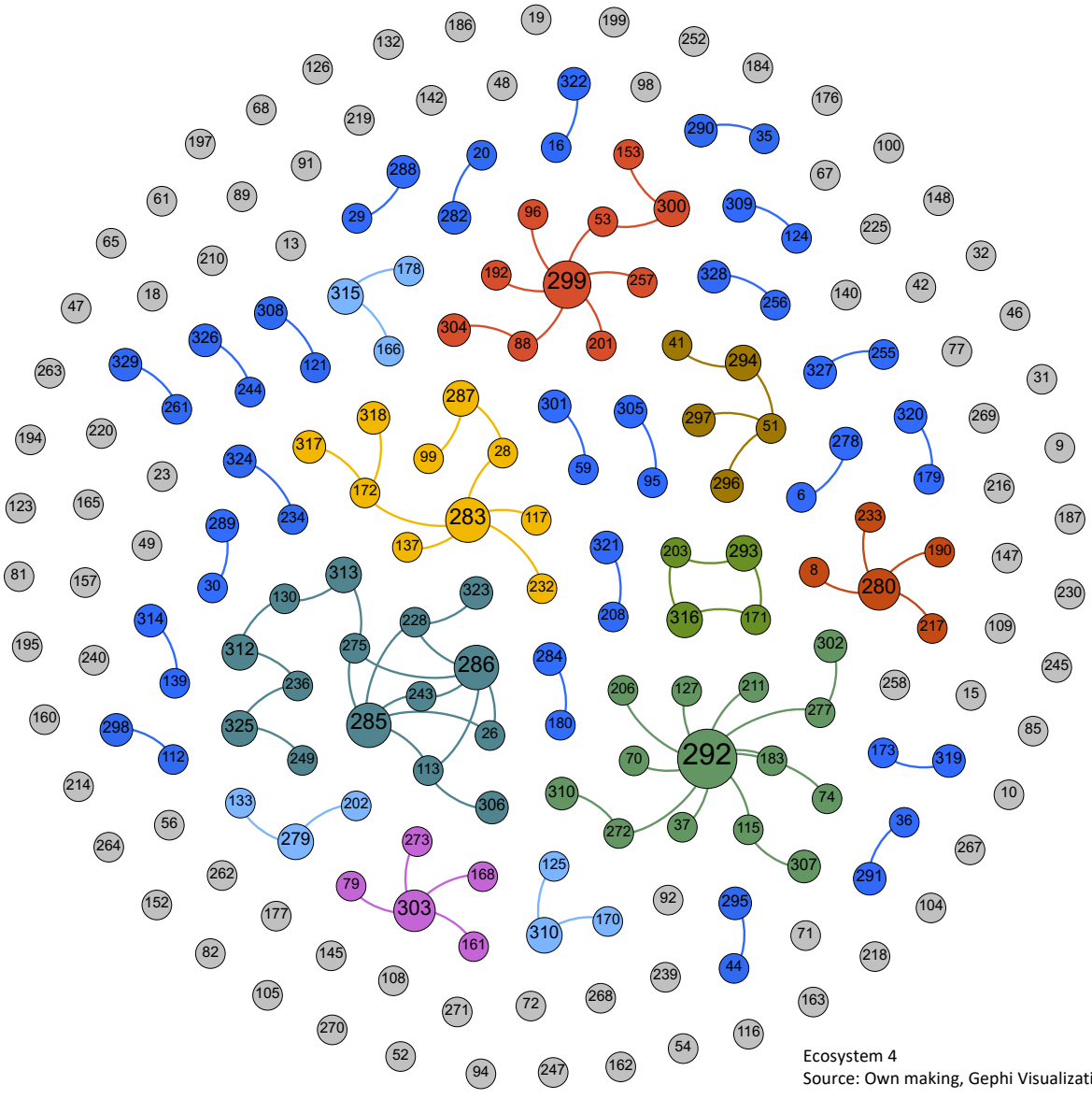


Ecosystem 3
Source: Own making, Gephi Visualizations

4.2.4 Clusters within the Ecosystem

The ecosystem visualized below is based on the same criteria as the ecosystem above. The difference between the two ecosystems is that we removed the top third-party domains with 10+ connections, to get a better understanding of the underlying patterns within the ecosystem. This allowed us, through the modularity formula, to create clusters within the ecosystem, which is visualized through the colours.

The modularity of colours shows a clear division of clusters. The small grey circles are pages without cookies, and therefore we did not find any clusters among these, the blue circles are pages with cookies from only one domain, etc. The modularity colour therefore showcases the division within the ecosystem. The ecosystem does not contain names, but instead IDs, to show the clusters and the connections more clearly. The clusters of relevance will be analyzed below.

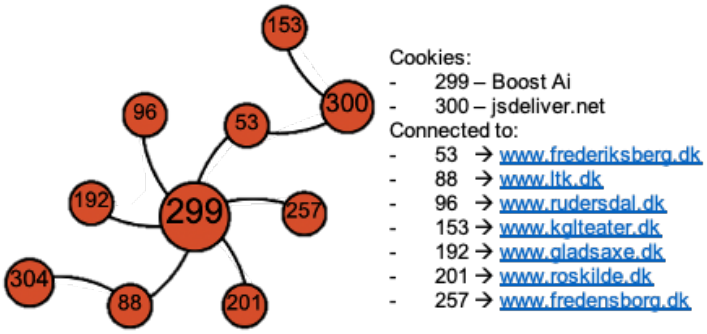


Ecosystem 4
Source: Own making, Gephi Visualizations

In the following section, the clusters with mainly pages on municipal level will be covered, which will then be followed by the clusters that contain pages on regional and national level respectively. The clusters with no relation between its pages will be covered in the end of this section.

Municipal Level

This cluster is based on the domain "*boost.ai*", which is owned by the company Boots.ai. Boost.ai only set cookies on these pages, which is a clear indication that only these pages are using the chat functions from Boost.ai. A visualization of the cluster and legend for the identification numbers can be seen below.



Cluster 6
Source: Own making, Gephi Visualizations

All of these pages are web pages of different Danish municipalities. All, but one, of these municipalities are located in the Copenhagen metropolitan area and in the area of North Zealand. Roskilde Municipality is not located in, but still near, the Copenhagen metropolitan area. Though there is not a clear depiction as to why only these municipalities use the services from Boost.ai, the region has established cooperation on digitalization. One example of cooperation in this region is Greater Copenhagen is a collaboration program with participation of the municipalities from Region Zealand and the Capital Region of Denmark, as well as municipalities from the regions of Skåne and Halland in Sweden. Within this collaboration program, the Greater Copenhagen Gigabit-project seeks to help the municipalities to learn from each other in order to strengthen digitalization (Greater Copenhagen, 2020). Whilst there is no mention of TPSs, one could argue that the Greater Copenhagen program would attend to it, since TPSs play a role in the digitalization of the public sector, as they are a part of the current public web page infrastructure.

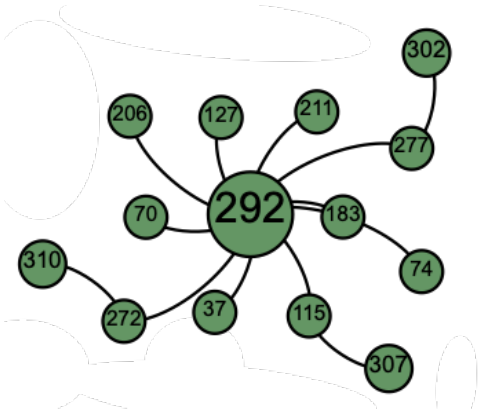
As mentioned in the analysis of cookies, see section 4.2.1, Boost.ai is Norwegian company that specializes in providing virtual agent and chat solutions to web pages of public administrations. Boost.ai provides its solutions through partners, such as consultancy firms, which are directly cooperating with the respective branch of a nation's public administration. This means that Boost.ai usage by the municipalities in this cluster, could be due to these municipalities working with Boost.ai partners.

Four of the six municipalities in this cluster, namely the municipalities of Frederiksberg, Lyngby-Taarbæk, Rudersdal and Gladsaxe, are members of Spar 5, which is a procurement partnership between five municipalities in the Copenhagen Metropolitan area. The last municipality in the partnership is Gentofte, which is not found in the cluster. The partnership is aimed towards ensuring cooperation regarding procurement, where it is possible between the affiliated municipalities (Frederiksberg, 2020). An advantage of this type of municipal procurement partnership, is that the municipalities can share knowledge among each other, which could be on the topic of digitalization. However, some municipalities may have to compromise and accept standardized solutions that might not be the ideal fit for the specific municipality (Kommunen, 2013). It is possible that the usage of Boost.ai's services is a result of these municipalities collaborating on procurement, especially since the cookies from Boost.ai have not been found on any other pages in the study.

This cluster is an example of geographically closely located municipalities that all use the same TPS. Though specific reasons for this are not determined, certain adjacently located municipalities in Denmark are cooperating in regard to procurement through partnerships, and especially in Region Zealand and the Capital Region of Denmark. These incentives could be a part of the reason as to why usage of Boost.ai as a TPS is restricted to the municipalities in the same region.

Regional Level

This section will cover two different clusters within the ecosystem due to their similarity. The first cluster is constituted by the web pages of hospitals located in, and thus governed by, the Capital Region of Denmark. The domain "*regionh.dk*", which is the domain of the Capital Region of Denmark, only set cookies on these pages. A visualization of the cluster and legend for the identification numbers can be seen below.

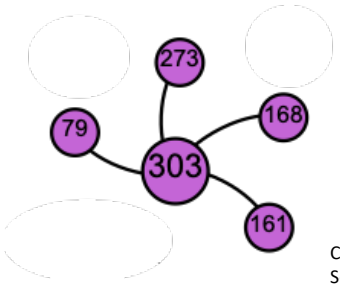


Cookie:
- 292 – Regionh.dk

Connected to:
- 37 → www.nordsjaellandshospital.dk/
- 70 → www.bispebjerghospital.dk
- 74 → www.rigshospitalet.dk
- 115 → www.amagerhospital.dk
- 127 → www.herlevhospital.dk
- 183 → www.hvideovrehospital.dk
- 206 → www.psykiatri-regionh.dk
- 211 → www.bornholmshospital.dk
- 272 → www.gentoftehosital.dk
- 277 → www.frederiksberghospital.dk

Cluster 7
Source: Own making, Gephi Visualizations

The second cluster is also constituted by pages for hospitals with cookies from their respective region. The cookie is set by the domain "*rm.dk*", which is the domain of the Central Denmark Region, and these hospitals are governed by this region. A visualization of the cluster and legend for the identification numbers can be seen below.



Cookie:
- 303 – rm.dk (Region Midtjylland

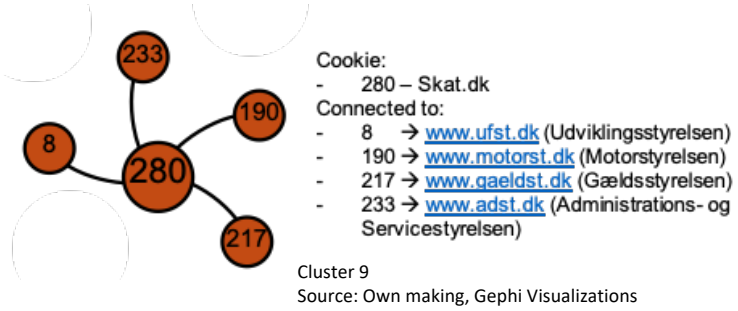
Connected to:
- 79 → www.regionshospitalet-horsens.dk
- 161 → www.auh.dk (Århus Universitetshospital)
- 168 → www.regionshospitalet-randers.dk
- 273 → www.hospitalsenhedmidt.dk

Cluster 8
Source: Own making, Gephi Visualizations

The names of the cookies from the domains of the two regions are different, and the names do not provide any indication of their purpose. The management of hospitals, including the psychiatric hospitals and departments, is governed and performed by the respective regions. According to Vrangbæk (2009) the regional operational organizations in Denmark, such as the hospitals, are characterized by a high degree of focus on efficiency and productivity. Working from this notion, it is possible that the cookies have been set on these pages with the purpose of optimizing the operations of these hospitals. However, Vrangbæk (2009) also describes that the regional authorities put larger emphasis on ethical awareness, compared to municipal and national authorities, which is presumably due to the fact that the regional authorities are operating the hospitals and thus have a direct influence on the health of the citizens. These claims from Vrangbæk (2009) may seem rather incompatible when it comes to setting cookies on web pages, but the diversity of cookies makes these clusters difficult to explain. One could argue that these cookies are ethically justifiable, since they have been set by the very branch of the public administration that operates the entities of the pages that the cookies were set on, and thus not any private TPSs. However, we do not know the exact reason nor purpose for these cookies, and their legitimacy are therefore up for further discussion.

National Level

This section will cover clusters, where the web pages are similar for national level entities. The first cluster is based on cookies from the domain "*skat.dk*" that were only set on four pages, which belong to four different agencies under control of the Ministry of Taxation. A visualization of the cluster and legend for the identification numbers can be seen below.

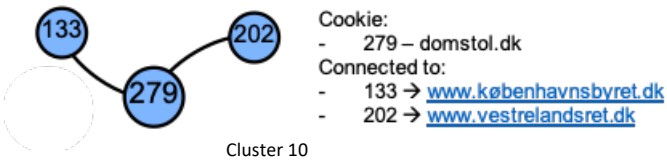


Cluster 9
Source: Own making, Gephi Visualizations

However, "*skat.dk*" is actually not the domain of the web page of the Ministry of Taxation nor the Taxation Authority (Skatteforvaltningen). "*Skat.dk*" is the domain of the web page that serves as the Taxation Authority's digital channel, where users can find the self-service access to registering taxes and guidebooks for doing it. As aforementioned, cookies from the "*skat.dk*" domain are only on pages from agencies under the Ministry of Taxation, such as the Danish Debt Collection Agency (Gældstyrelsen), which is in charge of collecting debts from individuals and businesses.

It is peculiar that these cookies have only been set on certain pages governed by the Taxation Authority. For example, the respective web pages for the Danish Property Assessment Agency (Vurderingsstyrelsen) and the Danish Customs Agency (Toldstyrelsen) did not contain any of these cookies. While there is a relevant connection between the implicated pages and the cookie domain, we found no plausible explanation as to why the domain "*skat.dk*" does not set cookies on the pages of all the agencies governed by the Ministry of Taxation

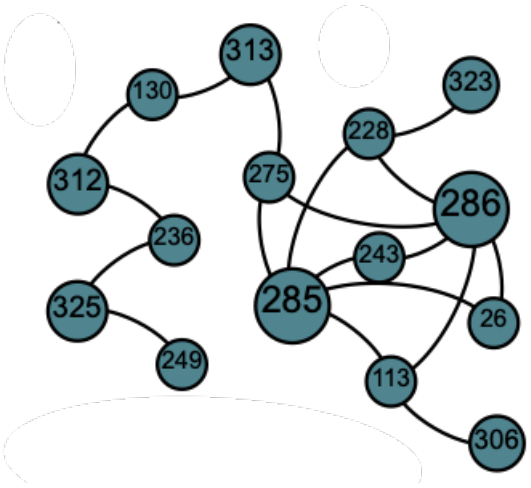
The second cluster is based on cookies from the domain "*domstol.dk*" that set cookies on pages for the legal courts in Denmark. This domain belongs to the Courts of Denmark, which are working independently from the Ministry of Justice, as they are located within two different branches of power. A visualization of the cluster and legend for the identification numbers can be seen below.



Cluster 10
Source: Own making, Gephi Visualizations

The cookies from this domain appear to be used for analytic purposes as they have names such as "*_ga*" and "*_gid*", which are known to be cookies that support the functionality of Google Analytics.

The last cluster on the national level differs from the two previously mentioned ones, as the cookie domain in this cluster is a private third-party domain. The domains "*youtube.com*" and "*doubleclick.com*", which are both owned by Google and thus the Alphabet Inc. conglomerate, only set cookies on five pages, where four of them are related to the Danish military. A visualization of the cluster and legend for the identification numbers can be seen below.



Cookies:
- 285 – youtube.com
- 286 – doubleclick.net
- 306 – soundcloud.com
- 313 – addthis.com
- 323 – ejendomsstyrelsen.dk
- 326 – pulseadnetwork.dk

Connected to:
- 26 → www.forsvaret.dk
- 113 → www.rmc.dk (rytmisk musikkonservatorium)
- 130 → www.nota.dk (Digitalt bibliotek)
- 228 → www.forsvaret.dk/fes
- 236 → www.dkpto.dk (Patent- og Varemærkestyrelsen)
- 243 → www.fmi.dk (Forsvarets material- og indkøbsstyrelse)
- 275 → www.hjv.dk (Hjemmeværnet)

Cluster 11
Source: Own making, Gephi Visualizations

Apart from the page domain "*rmc.dk*", which belongs to the Rhythmic Music Conservatorium, all the pages are related to the Danish military, such as the page domains "*forsvaret.dk*", which is for the Danish Defence, and "*hjv.dk*", which is for the Danish Home Guard. "*Youtube.com*" only sets three different types of cookies on these pages. As mentioned in section 4.2.1, these specific cookies are used to estimate the user's geographic location, bandwidth and previously watched YouTube videos (Cookiebot, 2020a). The domain "*doubleclick.net*" only sets one type of cookie, which is used for serving targeted advertisements to the user. The reason that these cookies are set, is likely that the clustered pages have embedded YouTube media plugins, which are used to display videos regarding the specific branch of the Danish military that the web page is used for.

Given the enormous scope of Google's operations, where it owns other domains, such as the prominent "*gstatic.com*", it is unlikely that Google should have a specific interest in setting cookies on web pages of the Danish military. However, this cluster could be an indication that the administration of Danish military favors using embedded YouTube video plugins to deliver content to the user, even though it enables Google to set cookies and thus retrieve information about the users. This could prove to be an ethical issue, as the web page for the Danish Defence holds

information that is relevant for a large group of people, such as information about the conscription. Thus, the user is forced to accept this TPT if he or she wishes to access this information.

Arbitrary Clusters

Some clusters appear to have a rather random or arbitrary page combination. Cookies from the domain "*facebook.com*" set cookies on the pages for entities such as the Danish Film Institute, Banedanmark, which is the governmental body of the Danish railway system, and Region South. None of these entities have any significant similarities in terms of focus area nor in terms of their roles in the public sector. This cluster is most likely just an indication that these pages all have a Facebook plugin installed.

The same pattern can be found within clusters based on cookies from the domain "*list-manage.com*", which is used by the email newsletter program called MailChimp. None of the implicated pages in each of these clusters have similarities, which are strong enough to suggest any kind of relationship.

4.2.5 Cookie Ecosystem Summary

In the findings and analysis of the clusters in the cookie ecosystem, we found two main patterns. Firstly, there is little to none relation between the web pages in some clusters. This is evident, since most clusters contain pages representing entities on different levels of government, i.e. municipal, regional and national levels, and thus indicates that the page combination has not been chosen selectively. Secondly, in some clusters the pages have some relation and these types of clusters exist with pages from all the three levels of government. In some cases, the domain that set the given cookie came from a private TPS, which are usually analytical tools and plugins. However, in other clusters the domain that set the cookie is identical to the domain of a public web page. In clusters based on these domains, the domains are most often identical to the page domains of the organizations that administer the entities that the pages in the cluster represent. In other words, governing entities and organizations set cookies on the pages of their entities, such as departments, agencies and hospitals. This makes it difficult to understand the true purpose of these cookies, however we can still see that they gather information about users.

4.3. Summary of Clusters in the Element- and Cookie Ecosystems

The two ecosystems are in general very similar, but we did find some noticeable differences. In terms of the general use of TPSs, the two ecosystems both exhibit a very arbitrary use of TPSs. However, the clusters in the element ecosystem seem to have higher level of arbitrariness than the clusters in the cookie ecosystem. Despite the arbitrariness, some clusters in both ecosystems do have pages that are interrelated, though this is more often the case in the cookie ecosystem. In certain cases, specific clusters from the element ecosystem are identical to clusters in the cookie ecosystem, indicating that both the cookies and elements have a functional relation. In regard to differences, the ecosystems differ in terms of pages that have an excessive use of TPSs. The element ecosystem contained clusters that are based on pages that make requests to multiple third-party domains, where the legitimacy of some of these domains is highly questionable.

Overall, the clusters in the cookie ecosystem are more structured, whereas the clusters in the element ecosystem seem to have far more arbitrary page combinations and TPS selections.

4.4 Stakeholder Analysis – Tracking, Trackers & Tracked

The issues in a public-private ecosystem, which in our case is an ecosystem about private/public tracking of public web pages, is the tensions between governments and the private sector, and the tension between national security, individual security and economic incentives (Raymond & DeNardis, 2015). We are witnessing several different stakeholders within the ecosystem and therefore contrasting incentives.

The stakeholders have different incentives and different kinds of power, legitimacy and urgency, as politicians have high power in enforcing laws, while organizations have high power based on resources. Another example is trackers with negative incentives, they have high urgency to be removed, while citizens have high urgency in relation to their protection of privacy. We are in this analysis providing an overview of the different stakeholders and their role within the ecosystem based on our conceptual framework and the preceding findings and analyses.

Stakeholders - Trackers

Top Third-Party Services (TPPS)

The TTPS are services from such as Google, Siteimprove and Twitter. These trackers are present on a great number of public web pages which is why they are top trackers. The extent seen within these trackers indicate they are a result of what Zuboff (2019) calls surveillance capitalism. They are earning money through tracking on an extensive part of the pages. This is where big data and the mosaic analogy for tracking becomes relevant, as the extent of data allows for identifying specific patterns about the users, which has become of source profit in targeting purposes (Libert & Nielsen, 2018; Boyd & Crawford, 2012)

These trackers have high power within the ecosystem according to the power definition by Mitchell et al., (1997), as they possess a great extent of resources and knowledge to be used for purposes such as targeting, automation and for profit. The legitimacy of these trackers is different depending on the purpose of the tracker. Trackers with the sole purpose of tracking, offering little in return to the web pages, such as TPS offering fonts, graphic elements, etc., have low legitimacy, as this makes no difference for the users, but users are tracked in return. TPSs that offer tools for assistance or analytics tools have higher legitimacy, as they help improve the pages by offering feedback. The TPPS are definitely high in urgency, as our findings showed that a great extent of the Danish public web pages use TPS with no real purpose, which in the end leads to “data leakage”, and this could be compromising personal data form Danish citizens.

Drawing on the theory by Blair and Whitehead (1988) about potential threat or cooperation from stakeholders within the ecosystem, trackers offering useful services such as analytics for web page improvement, are leaning towards being more of a collaborative partner, than an actual threat. TPSs who offer no real value to the user could be categorized as a threat, as they “leak” data without giving anything valuable in return (Blair & Whitehead, 1988). It is of course always a trade-off in the end, as users will lose some privacy at the cost of an improved service.

Small/Medium Third-Party Services (SMTPS)

Small/Medium Third-Party Services				
Power	Legitimacy	Urgency	Threat	Cooperative
Medium	Low-Medium	Medium-High	Medium-High	Low-Medium

The SMTPS resemble the top trackers as just described, however the difference is the extent of their tracking. The lesser extent of tracking by SMTPS indicates that they perhaps are not part of any commonly used tool. They do however still obtain information about the user without their consent. These TPSs are as the top trackers also profiting from tracking, as to why big data and surveillance capitalism once again is relevant (Libert & Nielsen, 2018; Boyd & Crawford, 2012)

These trackers do not collect the same amount of data as the top trackers, as to why their power is categorized as medium. The TPS we found within this segment are often smaller companies and do therefore not have the same resource power. They could however have stronger power in attaining more specific knowledge. The legitimacy of the SMTPS also differs between the specific trackers, as is the case with the TTPS. We did however find trackers that were either hidden behind proxies, hereby hiding their identity and corporate information, and trackers that were categorized as malware. These trackers have very low legitimacy as they could to a higher degree be dangerous to the privacy of the individual. Their low legitimacy therefore equals high urgency, due to the potential danger of these trackers. It is important that awareness is created about these trackers, to ensure better privacy of the individual.

The potential threat of these malicious trackers is therefore categorized as high, as they serve no real purpose for the site other than tracking and leaking data (Blair & Whitehead, 1988). The collaborative potential within the SMTPS is therefore low, as the services are not generally used within the ecosystem and the ones used could be harmful (Blair & Whitehead, 1998).

Public Third-Party Services (PTPS)

Public Third-Party Services				
Power	Legitimacy	Urgency	Threat	Cooperative
Medium	Medium	Medium	Medium	Medium

The PTPS is an interesting category, as this is public pages having tracking elements and/or cookies on other public pages within the ecosystem. Two examples are the pages with the domains "*regionh.dk*" and "*rm.dk*", who have tracking elements present on web pages for several hospitals in

their respective regions. The exact purpose of these trackers is difficult to pinpoint. A plausible explanation could however be a focus on efficiency and productivity, as a result of optimizing procedures within the hospitals. This raises the questions of why it is done through TPT, i.e. hidden tracking?

PTPS are categorized as having medium power, as they have limited resource power, but have higher political power. The purpose of these trackers is not to redistribute the data to other private organizations, but possibly to increase efficiency within the public sector. The legitimacy of PTPS is as well categorized as medium, as the exact purpose is unknown. The true legitimacy is dependent on the exact purpose. The legitimacy would be high if the sole purpose was to offer better services and experiences for the citizens using the web pages but would be lower if users do not gain anything relevant from the transaction. This stakeholder is categorized as medium in urgency, once again depending on the purpose of the tracking. The discussion of Lauritsen (2011) is relevant here, as the trust between people and government plays a great role in the Danish welfare model. The tracking conducted by public entities could therefore, once again, depending on the context of tracking be seen as a breach of that trust, as the majority of citizens are probably not aware that they are being tracked by their own government. This could therefore undermine the legitimacy of the state.

The undermining of legitimacy and hereby trust could be seen as a threat according to Blair & Whitehead (1988) as this tracking might not serve the interest of the citizens. It is therefore relevant to look into this kind of tracking and identify its true purpose. If the purpose is solely in the favor of improving the services for the citizens, the nature of the tracking would be more cooperative and not threatening to privacy to the same extent. It is however in the end a discussion of the trade-off between privacy and improved services.

Stakeholders: Users

Public Web pages (Politicians)

Public Web Pages (Politicians)		
Power	Legitimacy	Urgency
High	Low-Medium	High

The public web pages within the ecosystem consist of the page list we uploaded to the WebXray tool. These web pages have previously been described in more detail. The web pages are a product

of the digitalization and hereunder a product of the digitalization strategy by Danish government to achieve a higher level of e-government. A more digitalized state in the context of surveillance capitalism calls for high responsibility of the web pages to ensure the safety of its users, i.e. the citizens in Denmark. We did however find that several public web pages use third-party elements and cookies, which only have tracking purposes.

The power of the public web pages is categorized as high due to their political power, as they have the power to enforce through political initiatives once again returning to the definition of power by Mitchel et al., (1997). The web pages generally have the power to choose which services they wish to use, which in the context of our findings equals a legitimacy categorized as low to medium. Some of the public web pages we analyzed had an extensive amount of different TPS, which increases user data leakage. The use of TPS on Danish public web pages within the ecosystem shows some use of content that can arguably not be deemed necessary, as to why the legitimacy is categorized as low-medium. Some web pages do however seem to have better governance than others in the use of TPS. The general use of non-essential TPS therefore equals a high urgency to react on this stakeholder as data is leaking to several different TPS and might therefore be compromising the privacy of the individual (Mitchell et al., 1997).

General Users

The general users are those who visit the Danish public web pages, hereby being subjected to TPT. The use of public web pages is as previously stated a result of the focus on e-government in Denmark, as the citizens are “forced” to be online to gain access to certain public services. The transition towards a more digitalized state is however a result of an aggressive digitalization strategy as previously stated. The discussion by Lauritsen (2011) is relevant here, as the citizens rely on these web pages as part of their everyday lives, and when the safety is not optimal, it might be a breach of trust (Lauritsen, 2011).

The power and legitimacy of the general users are low, as they are the ones using the web pages, but they are simply lacking information about TPT even happening. There is little literature available in the field of TPT on public web pages and this issue is not on the political agenda in Denmark, meaning users are simply lacking knowledge about TPT on public web pages and the risk they are

subjected to. Low power and legitimacy are therefore a result of information asymmetry and digital illiteracy, as users do not have the power nor the knowledge to change anything. The asymmetric information deviation is further strengthened by complex IT-infrastructures. As a result of the information asymmetry and complex IT-infrastructures, Thakuriah et al., (2017) argues that digital illiteracy within digitized public sectors are increasing as well, i.e. users generally finding it harder to operate in a public digital environment based on complex private IT-infrastructures. All of the above therefore equals a high (very high) urgency of the general users, as they are being subjected to tracking, but generally have no knowledge due to the complexity and asymmetric information structure (Mitchell et al., 1997). This might in the end lead to a compromised fundamental right to privacy.

IT-Professionals

IT-professionals are also users of the public web pages. IT-professionals have an increased knowledge of the complexity of IT-infrastructures and the use of data. This increased knowledge allows for higher protection of being subjected to TPT, as they are aware of some kind of tracking going on (not necessarily TPT) and therefore uses plugins like ad-blockers and IP-disguises, such as VPNs (Virtual Private Network), which allows for “tunneling” your network through other IP-addresses, hereby hiding and protecting your own identity (Nordvpn, 2020).

The power of this stakeholder is still low, but higher than the general users, as IT-professional have the power to protect themselves against the tracking through increased knowledge. This stakeholder is not subjected to the same degree of digital/information illiteracy, and they can more easily cope within the complex ecosystem. They are however still subjected to information asymmetry, as they are still to some degree gaining nothing from using web pages, services, etc., from private trackers, but might be getting improved services from public trackers. They also have a little higher legitimacy, as the actions they are undertaking are proper and appropriate within the ecosystem, as they are protecting themselves and might help others to protect themselves. The legitimacy is evaluated as low-medium, as they have increased knowledge/power, but not sufficient to change anything crucial within the ecosystem. Just as general users, these stakeholders have high urgency as well, as we

argue that it should not be necessary to protect oneself from being tracked on Danish public web pages (Mitchell et al., 1997).

Tracking Experts

The last relevant identified stakeholder group is experts within the field of tracking. This stakeholder group is also subjected to tracking, but with even more knowledge than the IT-professionals, allowing them to protect themselves.

They do however differentiate themselves from the IT-professionals by having much higher power and legitimacy. These experts have extensive knowledge within the field of specifically tracking allowing them to be more proactive towards change. These experts have a voice in society, as to why their legitimacy is high. This voice could then be used to “push” the public and political debate about tracking and the extent hereof. Taking our findings into consideration about the extent of tracking on Danish public web pages, urgency is evaluated as high for experts, as they are obligated to voice their concerns surrounding this extent of tracking. It is an urgent need to ensure the protection of the fundamental right to privacy, as general users are probably not aware of being subjected to tracking.

Mikkel Flyverbom, Nanna Bonde Thylstrup, Rasmus Helles are examples of experts within the field of digitalization and tracking in Denmark, having published multiple articles and co-authored several pieces about tracking and its implications. Mikkel Flyverbom is now part of a newly established political branch called “Dataetisk Råd” (Data Ethics Council), with the purpose of ensuring responsible and sustainable data use (Digitaliseringsstyrelsen, 2020). Experts are essential to achieve a public debate about tracking, which is why it is important to study the field of tracking and once again relates to the importance and relevance of our thesis.

This concludes our findings and analysis section.

5. Theoretical Implications & Discussion

"...we can establish a critical sense that can help us set limits on headless digitization, uninhibited data harvesting and cynical automation. In this way, our public institutions can be based on values and ambitions that are our own and not driven by technological and commercial logic"

Mikkel Flyverbom (Flyverbom et al., 2019)

This section serves as a discussion of all the presented relevant concepts from the conceptual framework, findings and analyses, with the purpose of offering a broader contextual understanding of TPT within the ecosystem for Danish public web pages.

5.1 The Dilemma of the Digitalized State

In the findings and analysis section, we showed that user data is transmitted to various TPSs on the web pages of the public administration in Denmark. Given that this is the case, it is necessary to reassess the digitalization of the Danish public sector. As touched upon in the conceptual framework, the Agency for Digitisation states that citizens can expect public digital solutions to have the same quality standard as from private counterparts. However, according to Forrer et al., (2010), the complexity of an area such as digitalization does equal that the public sector is unlikely to achieve its goal without assistance from the private sector. This likely explains the usage of private TPSs having been covered in this thesis.

Even though we found user data transmitted to third parties, we also found that some pages are using their own servers to request certain elements. However, given that many of the third-party requests were made for dynamic content, such as special programs and plugins, it appears that the demand is for content that is created by more skilled private parties, making it difficult to replicate, once again returning to the lacking governmental technological competencies. Usage of TPSs thus seem inevitable, if the public web pages wish to use solutions such social media plugins, media players or analytical programs. These solutions may improve the usability, analyzability or cosmetic features of the pages, but might potentially be compromising user privacy. If citizens were to become aware that their online privacy is compromised on pages of the public sector as our findings show, it could potentially have a negative influence on the trust between the state and its citizens. Jensen & Svendsen (2009) argue that the citizens in Denmark believe in contributing to the common good, because they trust the government and public administration to ensure the country's welfare. Furthermore, Lauritsen (2011) claims that citizens accept that the government handles personal information, because the citizens trust the government to handle their data responsibly. That user data leaks to TPSs is hardly responsible data handling, especially considering that this user data is being transmitted without active acceptance of the user, i.e. consent.

On one hand, the inclusion of private parties is necessary to achieve the current desired level of digitalization at the desired pace, given that there is an information asymmetry between the public and the private sector, where the latter is ahead in terms of skills and knowledge. On the other hand,

user data leakage to third parties could potentially harm one of the social conditions for e-government, namely the citizens' trust in the government to handle their information responsibly. In other words, the usage of TPSs on Danish public pages could be both the enabler of and danger to e-government in Denmark, and thus the Danish government must evaluate the trade-off that is presented in this dilemma and reconsider whether it is possible to keep both TPSs and the people's trust over an extended period of time.

In our study, we also found that certain cases where regions set cookies on the pages of their hospitals, for example the domain "*regionh.dk*", which set cookies on pages for hospitals in the Capital Region of Denmark. Though it is still tracking, it seems to be different from cookies set by private TPSs, such as YouTube, because there might be a different agenda at play. What if the regions set cookies on the pages of their hospitals as a way to ensure better treatment of patients and thus enhancing health? Despite political and ideological views on how much the government should interfere with private matters, the TPT in this scenario could potentially benefit the user's health, even though it would technically still be compromising online privacy. This example illustrates the complexity of this topic and the many different factors at play, and that these factors should all be considered when re-evaluating the use of TPSs on Danish public web pages.

The information asymmetry between the public and private sector has led to a public IT-infrastructure, including internet pages, that has been partially built by private parties with their own interests. This is further concerning, given the monetary gains that some of the companies seek to achieve from the mechanics of surveillance capitalism.

5.2 Public User Data as a Source of Profit

As written in section 5.1, the concept of e-government is used in the Danish government as a digital channel to deliver public services and aid the Danish welfare system. Furthermore, e-government is made possible through the collaboration between public and private actors allowing for a more efficient public sector.

The Danish public it-strategy of being a leading player in operating a digitalized public sector, can only be realized by leveraging private competencies, as the Danish government has acknowledged that they do not possess the needed competences to develop or implement the necessary underlying IT-infrastructure. The strategy and implementation are an integral part of the Danish public sector

and the public institutions are therefore under political pressure to realize the strategy. This expected development does however lead to conflicting interests.

Private and public actors have different organizational incentives in running an organization. According to Omobowale et al. (2010), private and public actors have a conflict of interest, as public actors seek to improve efficiency, accessibility, usability, user experience, etc., which is consistent with the aforementioned digitalization strategy of the public sector. Private organizations on the other hand seek to improve profit, which can be done in several ways. One of the ways is collecting data to be used for improving processes or for resale to other organizations. Public institutions utilizing private IT-infrastructures to accommodate the political pressure might therefore lead to conflict.

It is clearly problematic that public authorities are forcing us to use services where information is collected externally (President of the IT-political association – Jesper Lund in Boye & Bredsdorff, 2017)

The different incentives lead to a conflict of interest and this might in the end make the citizens vulnerable and subjected to "data leakage" of personal information. Our findings and results from the WebXray output, data processing and the corresponding ecosystem visualization did indeed show an extensive use of TPSs on a large part of the Danish public web pages. This extensive use does all things equal mean that the citizens using Danish public web pages are being tracked, with great probability of the citizens not being aware of TPT. This is an issue, as the collected data from public web pages might be particularly sensitive information and therefore a breach of their fundamental right to privacy.

Tracking has become a source of profit, as the value of data has been increasing tremendously over the past decades. Private organizations tracking for profit have come to be known as Surveillance Capitalism according to Shoshana Zuboff (2019). Our findings are consistent with the concept of surveillance capitalism, as we found that a great part of the tracking elements and cookies serve no real purpose but to track and they come in several different forms, such as images, plugins, scripts, etc. The majority of the TPSs are not necessary for the web pages to operate. This raises the question of to what degree the Danish public web pages are actually aware of the tracking elements being embedded in the services they are using and if any kind of governance exists in the first place. We found no indication that any governance is enforced or is followed.

Surveillance capitalism differs from traditional capitalism, as traditional capitalism is based on supply and demand, whereas surveillance capitalism is based on endless accumulation. The endless accumulation of surveillance capitalism is also consistent with our findings, as the extensive amount of TPSs found are contributing to tracking. Every user entering the pages is being tracked without any specific purpose and is therefore inconsistent with the traditional model of supply and demand. Endless accumulation equals more data, which equals more information, which in the end allows for more precise predictions of the real world. This is made possible by big data analysis and algorithms according to Mayer-Schönberger & Cukier (2013) and Gillespie (2014). The combination of big data and algorithms for collecting and transforming data does however rely on human coding, which raises the question of objectivity and the use of the results. The results of a big data analysis should therefore not be perceived as objective data, as it is based on correlation rather than causality (Flyverbom & Madsen, 2015; Boyd & Crawford, 2012).

The primary issue of being tracked in general, but also through TPT, which is not visible to the user, is the lacking knowledge or information about the whereabouts of the data. This raises a lot of questions such as 'where does the data go', 'who has access to the data', 'who uses the data', 'are users subjected to influence without their knowledge', etc.? It is primarily the organizations tracking who are capable of answering these questions. Tracking is creating an imbalance in information between the parties, as to why information asymmetry is relevant in this context. We are arguing that information asymmetry combined with our findings showing extensive tracking of possible sensitive information, with an indication of no governance of the public web pages, is an issue.

In the end, tracking and the data as a result comes down to two polarized use-cases, i.e. usage vs. abusage. Usage vs abusage should be seen in the context of the citizen. Usage of data should equal improved services for the Danish citizens, as their data is collected. Abusage equals data collected, but nothing improved for the citizens, therefore abusage of the citizens personal data. Tracking can be used for the greater good, i.e. by improving services, efficiency, usability, research, etc. From our findings, we found that some of the public pages, e.g. "*regionh.dk*", tracked different hospitals, which are also public pages, as stated in section 5.1. Depending on the purpose of this tracking, this could be seen as a way of improving their services, e.g. people were searching for specific symptoms, etc., they would be able to offer better services within a specific area. If they however tracked with the purpose of reselling the data, it would no longer be usage, but abusage instead. In general, tracking with no real purpose on the pages such as fonts, images, and malware, could be seen as "abusage", as the citizens gain nothing from these trackers being present on the pages.

Even though data might be used for improving services, it still raises the question of the trade-off for the citizen. Is the balance in favour of protecting privacy or is it in favour getting improved services? In a study by Pew Research, an American nonpartisan think tank, it was found that the majority of the respondents described tracking as “creepy”, “Big Brother”, “Stalking”, etc., but it was also acknowledged that they were gaining benefits, as to why they chose to consent to tracking. The study-report was named “Free is a good price”, as it was found that people are willing to forfeit some privacy in exchange for benefits (Pew Research, 2016).

> “*…many Americans are willing to share personal information in exchange for tangible benefits, they are often cautious about disclosing their information and frequently unhappy about what happens to that information once companies have collected it.*” (Pew Research, 2016).

Consented personal information in exchange for benefits does however deviate from TPT, as TPT does not require consent. The lacking consent of TPT is most definitely an issue and should be addressed, which leads to the next section.

5.3 Challenged Fundamental Right to Privacy

If the issue of TPT on Danish public web pages is not addressed, the privacy of the individual in the shape of personal data might be subjected to misuse without the knowledge of the data subject. This should by all means be avoided to ensure the fundamental right to privacy according to Trzaskowski and Sørensen (2019).

The extent of tracking of today is a product of the commercialization of user activity, which is related to the concept of Surveillance Capitalism by Shoshana Zuboff. It is interesting to draw on Agre (1994) and his article about two models of privacy, which discusses the difference between “capture” and “surveillance”. Agre (1994) argued in his time of writing the article that capture and surveillance were two different ways of surveilling individuals. “Surveillance” should be seen as a political tool of control for catching misbehaviour, whereas “capture” should be seen as a tool for understanding human activity with the purpose of improving processes, and is therefore a logistical surveillance tool (Agre, 1994). The article by Agre is therefore just as relevant today, as it was back in the day, as the two modalities of “capture” and “surveillance” can be related to our current society.

Capture and Surveillance have today perhaps moved closer to each other as the “capture” of everyday activities is seen in an increasing number of areas, such as in retail stores, supermarkets,

stock markets, online user activities, on social media, etc., with the purpose of commoditization. The commercialization of personal activities online has come to be known as “surveillance capitalism” by Zuboff (2019). The traditional distinction of “surveillance” as a political tool and “capture” as a logistical tool by Agre (1994) is according to Zuboff (2019) therefore just as relevant, but the distinction have become blurrier, as “capture” of data is an essential part of surveillance capitalism. Surveillance is no longer solely a political tool, it is also a tool to “capture” data to be commoditized (Zuboff, 2019; Agre, 1994). The development since the writing of the Agre (1994) text has therefore been a commercialization of human activities, with the sole purpose of collecting information for not one specific purpose, but to create a bigger and more precise picture. This is consistent with the idea of TPT as a mosaic and is consistent with our findings, as we showed the extensive use of TPS within the ecosystem. We therefore argue that the right to privacy is under pressure within the Danish public sector.

This right to privacy is within the arena of TPT challenged by several factors, the primary one being “forced” to be online as a result of a partially digitalized public sector, i.e. e-government. We however found several underlying causes of why privacy might be compromised. The underlying factors are illustrated in the flow below:

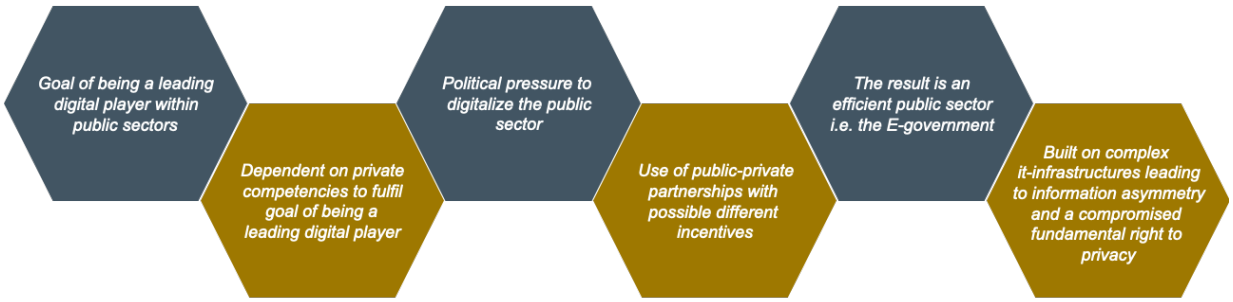


Figure 7: Politics within the public state
Source: Own making

The model illustrates the process causing the fundamental right to privacy being under pressure. It all starts with the digitalization strategy from the Agency of Digitisation with the purpose of being a leading player within digitalized public sectors. This digital development relies on strong competencies within technology and IT-infrastructure, which the Danish government has acknowledged it does not possess. The public institutions are expected to live up to the goal of being leading digitalized institutions, meaning they are under political pressure, which is why private organizations are leveraged to serve the purpose of being a leading player. This is however where conflicts of interest arise. Public and private organizations have different incentives, which in the end

might cause a compromised right to privacy of human subjects. But what does it actually mean to have a fundamental right to privacy? This right has developed over several decades from private photos being taken in hiding, to the atrocities of the world wars, and up till today where human activities are a source of profit. The fundamental right to privacy should be understood as the right to human dignity and other key values such as freedom of association and freedom of speech and is one of most important rights in the modern digital age (UN, 1948).

With the extent of tracking we are witnessing at the moment, and with our findings showing an extensive amount of tracking on Danish public web pages, we argue that privacy is under pressure within the Danish public sector. Even though some studies, e.g. Pew Research (2016), shows willingness to offer privacy for benefits, we argue that this should not be the case for the public sector, as it has another responsibility to protect its citizens due to the trust relationship previously described. We furthermore found little to no indication of governance directed towards decreasing or controlling the use of TPSs, as to why it further enhances the pressure of privacy due to lacking consent.

With the data leakage to TPSs that we have identified, it would be relevant to consider the future consequences. As mentioned earlier in the discussion about the Danish public sector's digitalization strategy, the usage of TPSs works as both the enabler of and danger to Danish e-government. Completely banning the use of TPSs on public internet pages would resolve the problem with this type of tracking, however it would also render it impossible to achieve the desired level of digitalization of the public sector. A more realistic approach would be to increase governance and create more specific guidelines for the use of TPSs on public pages. Guidelines for governance of the state by the state exist in other areas, such as government procurement. As mentioned in the analysis of the element ecosystem, certain pages requested video content from a SKI certified TPS. The SKI program could serve as inspiration for potential future certifications for TPSs. Though reducing the number of TPSs could lead to fewer companies gathering more user data, the public sector could still gain more control and in that way possibly avoid using illegitimate and illicit third-party domains like "*moatads.com*" and "*pulseadnetwork.com*". Furthermore, some pages request certain elements such as images and fonts from multiple different third-party domains, serving no real purpose, while other pages have solved this problem by making requests to their own servers, which would make it possible to avoid the use of TPSs.

While own servers and certification programs could be used for TPSs that provide necessary or essential content for the pages, such as analytical programs or elements that support disabled visitors, it is also important to reconsider the elements that can hardly be deemed essential. We found pages requesting third-party content such as social media plugins and media players, as well as other types of content that have cosmetic features. This type of content is not central to the purpose of the public pages, and thus its presence is hardly legitimate, considering that the price of it, is the transmission of user identifiable information to private companies that can be used towards the individual with advertisement purposes or worse.

According to Hempling (2014), policy makers are strongly influenced by powerful firms that operate in markets with a high concentration of power, which seems to apply to the market of TPSs with a giant like Google. Hempling's (2014) point could indicate that it will be difficult to push for increased governance, given that it would likely not be in the interest of the most powerful firms. On the other hand, government officials would likely not wish to endanger the people's trust that enables Danish e-government (Lauritsen, 2011) and thus essentially also parts of the foundation for the Danish welfare system (Jensen & Svendsen, 2009).

5.4 Technological Limitations, the "All-seeing eye" vs. Perfect Surveillance

Even though the privacy of individuals might be compromised, which of course is a negative aspect of our findings, it is relevant to widen the perspective of tracking and its limitations. The fear of data tracking is a very important issue in the digital age, as people are afraid of their privacy being compromised. Tracking is sometimes seen as the "All-seeing eye", due to the extent of tracking we are hearing about, e.g. NSA, Google, Apple and Facebook. There is no doubt that a lot of data is flowing between systems on a daily basis, as we previously wrote that the world is producing 5,8 gigabytes of data each day, but the quality of this data and tracking data in particular is up for discussion, which deviates from data being an "all-seeing eye".

The majority of technology, hereunder tracking technology, is very basic and built by humans, which means they have limitations and it therefore limits the output to human errors and the capabilities of humans. No technology is perfect, as to why the tracking of today has still not reached the level of a dystopian surveillance society where nothing goes unnoticed. Fear of tracking is not unreasonable, but the quality is however up for discussion. It is very hard to precisely predict the individuals who are tracked, their real intentions, their social aspects, their real information needs, etc. (Fourie & Bothma, 2007). An example is the facial and affect recognition industry, which is about assessing

patient pain, customer emotion, student attentiveness, etc. This industry is undergoing significant growth according to Sawers (2019). The demand for these technologies is surging, as organizations want to be on the forefront and be able to know possible candidates and possible customers as detailed as possible (Crawford et al., 2019). The technology does however have its limitations.

> *"There remains little to no evidence that these new affect-recognition products have any scientific validity"* (Crawford et al., 2019:51)

Perfect surveillance would be surveillance capable of catching perpetrators of crimes, without surveilling individuals with no relation to the crime. This scenario is however a utopian scenario, which is extremely hard to reach, as algorithms would have to be more complex than we could imagine (Re, 2016). Perfect surveillance is probably not possible in the near future. Tracking is very complex and based on several underlying factors, as to why there is no simple explanation or solution. The suboptimal solution of tracking would be to be able to map out dataflows more precisely and ensure clear governance, hereby being able to always answer questions such as "how does the flow look", "where does the data go", "who has access to the data", "what is the use-case of the data", etc.

All things said, it is still very important to be critical towards data, as data might be biased, or subject to errors. An example of this is the study conducted by Latanya Sweeny (2013), which found that a user was more likely to be shown advertisements related to criminal records databases when searching for popular African American names, compared to popular Caucasian names, and other cases as previously written. Technologies might be flawed and prone to errors, which is why critical use is essential.

> "*Facial expression recognition technology is picking up on something — it's just not very well correlated with what people want to use it for. So they're just going to be making errors, and in some cases those errors cause harm,"* (Professor Jonathan Gratch in Crawford et al., 2019:51)

In the context of TPT on Danish public web pages, we found that the Danish Gambling Authority had a lot of TPSs, which equals more tracking of individuals on this web page, which in the end might lead to more specific targeting towards this group. Users who visit the web page of the Danish Gambling Authority might be a group who is vulnerable in the first place.

6. Conclusion

"We are seeing that at many political and private leaders a feeling exists that technologies are a train that is leaving without us unless we jump on it. But the question is where the train is going and whether it is the destination we wanted to go to. Time is rarely given for reflection over this"

Nanna Bonde Thylstrup (Persz, 2020)

The purpose of our thesis is to explore the current ecosystem of TPT on Danish public web pages and to explore the underlying incentives such as politics, efficiency, profit. The purpose is to establish a better understanding of the current landscape of tracking. To achieve this purpose, we found a list of web pages belonging to the Danish public administration (national focus) and used WebXray to analyze these pages for third-party trackers (tracking focus). The output was then transformed, analyzed and put into relevant contexts for discussion.

Our findings of the third-party ecosystem showed that Danish public web pages use various TPSs and that TPT exists on the majority of the analyzed pages. We found that several different requested file types, e.g. images and fonts, with JavaScript code elements being the by far the most requested element type.

We found several different purposes of the elements within the ecosystem. Some of the requested elements appeared to be visible to the users, such as images, fonts and other cosmetic features. Other elements appeared to be used only by the web page owners or developers, such as analytical tools. Some TPSs received requests for more unusual elements, e.g. an image element from Google in the shape of an 1x1 pixel in size and only used for tracking purposes, and the "*trackingCode.js*" element from the font provider "*fonts.com*". Third-party element requests are highly concentrated around a few large TPSs, such as Google, Twitter and Siteimprove hereby illustrating a long-tail divide. We argue that the need for the majority of the elements is questionable, especially considering that they cause a data leakage. Furthermore, the legitimacy of some of these less common TPSs can also be questioned, because they intentionally hide company information and are occasionally strongly linked with malware and illicit content. Third-party cookies follow a similar pattern as elements. However, cookies seem to be used more specifically in relation to certain tools and plugins, such as analytics programs and media players. Google and Siteimprove are the private TPSs that have third-party cookies on most pages. However, like with the third-party elements, some TPSs that set cookies have questionable legitimacy, as they are linked to malware.

In terms of the ecosystems, the cookie ecosystems contained clusters with closer relations between the TPSs and pages than compared to clusters in the element ecosystems. Furthermore, some public entities set cookies on the pages of entities under their control. This "internal" tracking is also the case with regions setting cookies on pages for hospitals. Despites some differences between the two ecosystems, the overall impression is that there is a high level of arbitrariness in terms of which TPSs are used by which pages.

Drawing upon our findings, several aspects have become evident. Tracking has become a source of profit, which is consistent with the concept of Surveillance Capitalism and our findings, as we found clear evidence that user data is transmitted to several TPSs.

We argue that there are multiple causes of the extensive leak of data to TPS on Danish public web pages. One such cause is that the Agency of Digitisation is pushing the agenda of being a leading digitalized public sector, promising public solutions on par with quality seen within private solutions. This demands strong private competences, as the Danish government has acknowledged it does not possess these competences. The use of private resources such as IT-infrastructures, services, etc., does however lead to a conflict of interest as previously described. To reach the desired public digitalization level, use of TPSs is inevitable, it therefore comes down to the trade-off between usability and efficiency, and the privacy of individuals.

We argue that the use of private resources in the digitalization of the public sector is in conflict with the privacy of individuals, as Danish citizens are “forced” to be online. Citizens are therefore subjected to tracking without their consent and are therefore vulnerable to tracking of sensitive data, which in the end leads to the fundamental right to privacy being under pressure. The government needs to reevaluate, if the desired level of digitalization and current level of trust can persist concurrently, or if one or the other has to be dominant.

As privacy is under pressure within the Danish public sector and that we found very little indication of governance or codes of practice within the third-party ecosystem, we argue that a reassessment of public digitalization strategy is necessary, for the trust towards the Danish state to be sustained and for the protection of the fundamental right to privacy. Without any action, trust could be influenced negatively, hereby damaging the essential basis of the Danish welfare system. A lot of trust in the form of data is placed in Danish public institutions and in return, citizens expect responsible handling of their data. Our findings of extensive leaking can however not be considered as responsible handling.

The reassessment of the digitalization strategy demands governance and codes of practice with the purpose of ensuring a more responsible use of TPSs, where the necessity of elements is evaluated more thoroughly. Technologies should be central for the existence of the page and not serve as a supplement, e.g. social media share buttons. The Danish government should evaluate the trade-off

between usage and abusage in the context of citizens, i.e. if citizens are gaining something or if their data is abused for other purposes, thereby compromising their privacy. It is therefore crucial to map the whereabouts of the data in order to ensure proper protection, asking the questions of who is collecting, who has access, where the data goes, etc., to identify the legitimate interest of the tracking and if it should be interrupted.

The relevance of our findings in the bigger picture illustrates numerous aspects. Tracking has become an industry, which is why we see an increase in tracking and the extent of tracking happening at the moment. We should without a doubt be critical towards the extent of tracking, as individual privacy is under pressure. This is especially evident in this study as some pages are more prone to contain sensitive information than others. Tracking is sometimes perceived as an "all-seeing eye", but the fact is that technology is not perfect. The majority of technology is prone to errors, as it is limited to human capabilities, which means the data quality is not perfect and therefore not as dangerous as initially perceived, but still calls for critical use. The use of technology should not be seen as a train leaving hereby missing out on technological advancements, but as a legitimate need to improve society in general, based on clear governance, guidelines and regulation.

6.1 Further perspectives: COVID-19, Surveillance and Citizen Privacy

The right to privacy of data subjects has been a central topic throughout our thesis, as we through our findings and corresponding analysis have shown that the Danish public sector is subjected to an extensive amount of TPT. We wish to shortly broaden the perspective of our thesis as we at the time of writing are in the midst of the COVID-19-crisis, which allows for drawing on peripheral, but interesting perspectives with relevance to our study.

We therefore in this section discuss COVID-19 and its implications for the digitalized state of Denmark. When a crisis hits, it is often unexpected and therefore defining for our current society, demanding quick decisions. These rapid reactions and decisions might however be the cause of changing future societal dynamics. 9/11 is an example of a crisis leading to changing dynamics. In the repercussions of 9/11 a shift was seen from privacy and democratic rights towards state security in the United States. NSA was established shortly after and the giant tech companies entered into hidden partnerships with intelligence agencies, hereby canceling the initial thoughts of protecting privacy of the individual (Lyon, 2001, Flyverbom, 2020). It is argued by Shoshana Zuboff that the digital surveillance as a result of "national security" has caused an unregulated field of commercial tracking, which is known as surveillance capitalism (Zuboff, 2019).

The corona crisis is a different crisis but calls for the same considerations. Fear and insecurity should not be used as tool for implementing further surveillance, which would not have been accepted in the first place. The immediate choice of the Danish state would have been to monitor the Danes if they were staying home, grouping, etc.

"*Governments around the world are in the process of implementing analogue as well as highly technological solutions, with the purpose of containing the infection*" (Persz, 2020)

At least 25 countries are planning to implement collection of data from mobile devices, apps, etc., which might compromise privacy (Persz, 2020). The Danish state and a big part of the Western world did however choose not to monitor digitally through arguments of safety and privacy. Google sought to implement a "Corona-app" allowing for tracking of the infection to be mapped all over the world, which would have equaled more data given away. Due to stronger privacy guidelines of Denmark, the Agency of Digitisation denied the app to be implemented in Denmark, as a similar Danish app existed using Bluetooth instead and is therefore not tracking geolocation data. This is interesting in relation to our findings, which are however not related to any crisis, but still to a possibly compromised privacy, TPSs, hereunder TPT, are already existing and to the fact that we found no indication of governance. The arguments of safety and privacy are therefore contradicting our findings.

The question raised by Flyverbom (2020) in the context of a crisis is "*If technological solutions can help us out of the worst crisis in history, should we then use them at any cost"?* Several professors, such as Mikkel Flyverbom, Professor & member of the Data Ethics Council of Denmark, Stine Bosse, head of the TechDK Commission, Nanna Bonde Thylstrup, PHD and Lector in Communication and Digital Media, argue that we should be careful in implementing new technologies too rapidly, as it might compromise privacy (Persz, 2020 & Flyverbom, 2020)

The TechDK commission has composed guidelines for helping navigate within the Corona-crisis thereby ensuring privacy, democracy and ensuring that tech organizations are not gaining more power at the cost of citizen privacy, when citizens are vulnerable. Stine Bosse argues in Persz (2020) that we should not be afraid to discuss the downside to surveillance in times like these, as she argues that we need transparency about where data is stored, who controls it and the legal background for collection and use.

The issue with introducing new technologies to accommodate a crisis as the corona-crisis, is what Nanna Bonde Thylstrup calls “function creep”. Function creep is a tendency, where technologies are developed for one purpose, but used in other contexts. Nanna Bonde Thylstrup further argues that governmental and private institutions show great interest in the development of these technologies, which could be problematic. “*When such actors come into play, and often in collaboration with government institutions, there is a real risk of abuse in the form of function creep*” Nanna Bonde Thylstrup in Perz (2020). This is consistent with the theory of Omobowale et al. (2010) arguing that private-public conflict often is a result of different incentives, i.e. the public need for competences to be more efficient vs. the private incentive of increased profit.

It is further argued by Nanna Bonde Thylstrup that Denmark is lacking a proper democratic discussion of how technologies are to be used and that we rarely have such discussions in Denmark. She argues that this is a result of not being “raised” to think like that, as we do have a history of trust to the state, whereas in Germany, a history of mistrust is more evident, having prepared them better for these situations (Persz, 2020). Trust to the state in Denmark could and should therefore be treated with care.

The Danish response is a reflection of digital responsibility, and a response towards tech companies and the complexity of these organizations managing critical societal functions. It is however interesting to view the stand and actions of the Agency of Digitisation in relation to our findings, as our findings are contradicting the stronger privacy guidelines of not being subjected to tracking. We did show extensive TPT on a great number of public pages and found little to no indication of any governance of the use of TPSs.

Privacy, freedom and trust to the state must be sustained, in order to keep running society with its current model, which is that of well-functioning e-government. It is therefore essential that we embrace technology at our own pace and do not view technology as a train leaving without us. We should not adopt technologies that in the moment seems evident, without thinking about the future consequences. The lacking governance in our findings could therefore be a result of adopting services and technologies too rapidly, which in the end might compromise privacy.

6.2 Limitations & Implications of the Study

Given that we study usage of TPSs on Danish public internet pages at the time of this thesis, this study could only be replicated within a short period of time or not be replicated at all. Though the exact same methodology could be applied, the usage of TPSs on the internet can be modified continuously based on new needs and regulation, which makes it difficult to replicate the study. Thus, an attempt to replicate the study would likely not yield the same results, however it would add knowledge to our understanding of the topic.

Given our academic background in business and communication, we have used WebXray as a tool for the data collection, as it is accessible for scholars in branches of academia other than computer science. This also means that certain aspects of tracking of TPSs on internet pages are too technologically demanding for us to study it. Furthermore, given that we have centered our empirical data collection around WebXray, we are also vulnerable for any shortcomings that the program may have had. Though there is no reason to expect that WebXray is faulty in any way, given that it has been used in numerous other studies, including Libert (2015) and Helles et al. (2019), we must still consider this as a factor, because we rely on its output in this study.

Another limitation for the study is the selection of pages. We included the internet pages of public entities listed in the Agency for Digitisation's overview of the Danish public sector, where we also added the respective internet pages for the hospitals in Denmark. Though we believe that this covers a broad part of the Danish public sector, pages for public schools, universities and current campaigns were not included. Furthermore, pages for state-owned enterprises, such as Ørsted, were also not included. Including these pages could show a broader ecosystem of third-party trackers and could be relevant to include in any future studies on this topic. Since, we the authors, are business and communication scholars, it would be relevant to get more perspectives on this topic and that could be achieved by studying it in other academic disciplines. As aforementioned, scholars in data science would be able to study the technical aspects of TPT, which we are unable to study, as we lack the necessary skills.

Another interesting academic discipline in this case is law. Law scholars would be able to study how TPT should be considered from a legal point of view, which could be relevant in order to better understand the necessary regulation and governance on this area. Studying the topic of TPT from different perspectives, would add knowledge and essentially help to assess the legitimacy and size of the problem that TPT poses to privacy.

Reference List

Acker, S.V., Hausknecht, D., & Sabelfeld, A. (2017). Measuring login webpage security. SAC '17.

Adgang med tegn. (2020). Adgang med tegn. Adgangmedtegn.dk. Retrieved 28 April 2020, from http://adgangmedtegn.dk.

Adobe. (2020). Audience Manager cookies. Retrieved 28 April 2020, from https://docs.adobe.com/content/help/en/core-services/interface/ec-cookies/cookies-am.html

Agre, P. E. (1994) Surveillance and capture: Two models of privacy, The Information Society, 10:2, 101-127, DOI: 10.1080/01972243.1994.9960162

Alteryx. (2020). About Us. Alteryx. Retrieved 16 April 2020, from https://www.alteryx.com/company/about-us.

Andersen, L. B., Hansen, K. M., & Klemmensen, R. (red.) (2012). Metoder i statskundskab. (2 udg.) Kbh. Samfundsvidenskabernes metoder, Nr. 1, Bind. 1.

BD. (2020). What is information asymmetry? definition and meaning. BusinessDictionary.com. Retrieved 8 May 2020, from http://www.businessdictionary.com/definition/information-asymmetry.html.

Blair, D. L., & Whitehead, C. J. (1988). Too many on the seesaw: Stakeholder diagnosis and management for hospitals. Hospital and Health Administration, vol. 33, pp. 153-166

Bloomberg, J. (2018). Digitization, Digitalization, And Digital Transformation: Confuse Them At Your Peril. Forbes. Retrieved 9 March 2020, from https://www.forbes.com/sites/jasonbloomberg/2018/04/29/digitization-digitalization-and-digital-transformation-confuse-them-at-your-peril/#16de49042f2c.

Boost.ai. (2020). Conversational AI for government and public sector. Retrieved 28 April 2020, from https://www.boost.ai/conversational-ai-public-sector

Booth, W. C., Colomb, G. C., & Williams, J. M. (2008). The craft of research (3rd ed.). Chicago: The University of Chicago Press.

Boyd, D., & Crawford, K. (2012). Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. Information, Communication, & Society, 15, 662-679.

Boye, M. & Bredsdorff, M. (2017). Offentlige hjemmesider deler følsomme data om danskernes netbrug. Teknologiens Mediehus. Retrieved from https://www.version2.dk/artikel/offentlige-hjemmesider-deler-foelsomme-data-danskernes-netbrug-1075728

Bucy, E. P., & Zelenkauskaite, A. (2016). A scholarly divide: Social media, Big Data, and unattainable scholarship. First Monday, 21(5)

Cadwalladr, C. (2014). Are the robots about to rise? Google's new director of engineering thinks so…. the Guardian. Retrieved 26 February 2020, from

https://www.theguardian.com/technology/2014/feb/22/robots-google-ray-kurzweil-terminator-singularity-artificial-intelligence.

Caulfield, M. (2017). Web literacy for student fact-checkers. Textbooks 5.

Ciborra, C. (2006). Imbrication of Representations: Risk and Digital Technologies. Journal of Information Management Studies, Vol. 43, No.6: pp. 1339-1356.

Cleanpcsolutions.com. (2016). How to get rid of Pulseadnetwork.com. Cleanpcsolutions.com. Retrieved from https://www.cleanpcsolutions.com/tag/how-to-get-rid-of-pulseadnetwork-com/.

Cludo. (2020). Get more out of your content. Cludo. Retrieved 11 May 2020, from https://www.cludo.com/

Computer Fraud & Security. (1999). UK public ready for E-government, 1999(7), 4.

Confirmit. (2020). Insight Software Solutions for VoC, VoE & Market Research Programs. Confirmit. Retrieved 12 May 2020 from https://www.confirmit.com/What-We-Do/?feed=20863852-ff99-4b4f-9762-f37b601a3c7a

Cookiebot. (2020a). Cookiedeklaration. Retrieved 28 April 2020, from https://www.cookiebot.com/da/cookie-declaration/

Cookiebot. (2020b). GDPR and cookies | GDPR cookie consent | Is my use of cookies compliant?. Cookiebot.com. Retrieved 8 May 2020, from https://www.cookiebot.com/en/gdpr-cookies/.

Cookiepedia. (2020). _ga cookie name search results. Cookiepedia.co.uk. Retrieved 28 April 2020, from https://cookiepedia.co.uk/cookies/_ga.

Crawford, K., Dobbe, R., Dryer, T., Fried, G., Green, Kaziunas. E., Kak, A., Mathur, V., McElroy, E., Sánchez, A.N., Raji, D., Rankin, J.L., Richardson, R., Schultz, J., West, S.M., & Whittaker, M. (2019). AI Now 2019 Report. New York: AI Now Institute. Retrieved from https://ainowinstitute.org/AI_Now_2019_Report.html.

Datatilsynet. (2018). Vejledning om håndtering af brud på persondatasikkerheden. Datatilsynet. Retrieved from https://www.datatilsynet.dk/media/6558/haandtering-af-brud-paa-persondatasikkerheden.pdf.

Datatilsynet. (2020). DMIs behandling af personoplysninger om hjemmesidebesøgende. Datatilsynet.dk. Retrieved 28 February 2020, from https://www.datatilsynet.dk/tilsyn-og-afgoerelser/afgoerelser/2020/feb/dmis-behandling-af-personoplysninger-om-hjemmesidebesoegende/.

Davies, H. (2015). Ted Cruz using the firm that harvested data on millions of unwitting Facebook users. The Guardian. Retrieved from https://www.theguardian.com/usnews/2015/dec/11/senator-ted-cruz-president-campaign-facebook-user-data

Defgo. (2020). Defgo spørgeskemasystem, trivselsundersøgelse og APV. Defgo.com. Retrieved 28 April 2020, from https://www.defgo.com/dk/.

Digitaliseringsstyrelsen. (2016). A Stronger and More Secure Digital Denmark: The Digital strategy 2016-2020. Copenhagen: Digitaliseringsstyrelsen.

Digitaliseringsstyrelsen. (2018). Denmark leads the world in e-government. Retrieved from https://en.digst.dk/news/news-archive/2018/september/denmark-leads-the-world-in-e-government/

Digitaliseringsstyrelsen. (2019). Det offentlige Danmark 2019: Oversigt over indretningen af den offentlige sektor. Finansministeriet.

Digitaliseringsstyrelsen. (2020). About the Agency for Digitisation. En.digst.dk. Retrieved 28 February 2020, from https://en.digst.dk/about-us/.

Dodds, L. (2017). What is data asymmetry?. Lost Boy. Retrieved 17 March 2020, from https://blog.ldodds.com/2017/03/24/what-is-data-asymmetry/.

Domains By Proxy. (2020). Domains By Proxy. Domainsbyproxy.com. Retrieved 28 April 2020, from https://www.domainsbyproxy.com.

Dresch, A., Lacerda, D., & Antunes, J. (2015). Design science research: A Method for Science and Technology Advancement. Springer.

Dubé, L., & Paré, G. (2003). Rigor in information systems positivist case research: current practices, tre. MIS Quarterly, 27(4), 597–635.

Duffy, D. (2019). How to Deal with the Challenges of Digital Law in 2019. Equities News. Retrieved 17 March 2020, from https://www.equities.com/news/how-to-deal-with-the-challenges-of-digital-law-in-2019.

Eisenhardt, K. M. (1989). Building theories from case study research. Academy of Management Review, 14(4), 532–550.

Ellram, L. M. (1996). The use of the case study method misconceptions related to the use. Journal of Business Logistics, 17(2), 93–138.

Erhvervsstyrelsen. (2017). Retningslinjer for cookies på hjemmesider for offentlige myndigheder. Erhvervsstyrelsen.dk. Retrieved 3 March 2020, from https://erhvervsstyrelsen.dk/vejledning-retningslinjer-cookies-paa-hjemmesider-offentlige-myndigheder.

European Parliament. (2017). Data Flows - Future Science. Directorate General for Internal Policies, Policy Department A: Economic and Scientific Policy.

European Commission. (2019). Web Accessibility - Shaping Europe’s digital future. European Commission. Retrieved from https://ec.europa.eu/digital-single-market/en/web-accessibility.

Extreme Tracking. (2020). eXTReMe Tracking. Extremetracking.com. Retrieved 11 May 2020, from https://extremetracking.com.

Facebook. (2020). Cookie Policy. Retrieved 28 April 2020, from https://www.facebook.com/policy/cookies/

Finansministeriet. (2019). Svar på Erhvervsudvalgets spørgsmål nr. 18 af 26. september 2019 stillet efter ønske fra Lisbeth Poulsen (SF). Finansministeriet. Retrieved from https://www.ft.dk/samling/20182/almdel/eru/spm/18/svar/1600557/2094130/index.htm

Flyverbom, M. (2019). Debat: Overvågningskapitalisternes mål er at kontrollere vores adfærd. Information. Retrieved 4 May 2020, from https://www.information.dk/debat/2019/04/overvaagningskapitalisternes-maal-kontrollere-vores-adfaerd.

Flyverbom, M. (2020). Tech-klumme: Lad os ikke gentage fejlene fra 9/11 under coronakrisen. Politiken. Retrieved 30 April 2020, from https://politiken.dk/viden/Viden_og_Tech_analyser/art7761564/Lad-os-ikke-gentage-fejlene-fra-911-under-coronakrisen.

Flyverbom, M. & Madsen, A.K. (2015): Sorting data out – unpacking big data value chains and algorithmic knowledge production, in Society of Data. Transcript Verlag.

Fonts.com. (2020). About us. Monotype. Retrieved 12 May 2020 from https://www.fonts.com/info/about-us

Forrer, J., Kee, J., Newcomer, K., & Boyer, E. (2010). Public–Private Partnerships and the Public Accountability Question. Public Administration Review, 70(3), 475-484.

Fourie, I., & Bothma, T. (2007). Information seeking: an overview of web tracking and the criteria for tracking software. Aslib Proceedings, 59(3), 264-284. https://doi.org/10.1108/00012530710752052

Franzke, A. S., Bechmann, A., Zimmer, M., Ess, C., & the Association of Internet Researchers. (2020). Internet Research: Ethical Guidelines 3.0. Retrieved from https://aoir.org/reports/ethics3.pdf

Frederiksberg. (2020). Om udbud og indkøb. Retrieved 28 April 2020, from https://www.frederiksberg.dk/virksomhed/udbud-og-indkob/om-udbud-og-indkob

Freeman, R. E. (1984). Strategic management: A stakeholder approach. Boston: Pitman.

Gephi. (2020). The Open Graph Viz Platform. Gephi. Retrieved 12 May 2020 from https://gephi.org

Gillespie, T. (2014). The Relevance of Algorithms, In Media Technologies, ed. Tarleton Gillespie, Pablo Boczkowski, and Kirsten Foot. Cambridge, MA: MIT Press.

Google Analytics. (2020). Analytics Help. Support.google.com. Retrieved 11 May 2020, from https://support.google.com/analytics/#topic=3544906.

Greater Copenhagen. (2020). Greater Copenhagen Gigabit - et signaturprojekt. Retrieved 28 April 2020, from https://www.greatercph.dk/projekter/gigabit

Greve, B. (2018). At the heart of the Nordic occupational welfare model: Occupational welfare trajectories in Sweden and Denmark. Social Policy & Administration, 52(2), 508-518.

Greve, C., Lægreid, P., & Rykkja, L. H. (red.) (2016). Nordic Administrative Reforms: Lessons for Public Management. London: Palgrave Macmillan. Public Sector Organizations https://doi.org/10.1057/978-1-137-56363-7

Helles, R., Lomborg, S., & Lai, S. S. (2019). *The Invisible Internet: Mapping of TPSs as a new resource for analyzing and comparing digital media systems*. Abstract fra Comparative Media Studies in the Digital Age, Beijing, Kina.

Hempling, S. (2014). "Regulatory capture": sources and solutions. Emory Corporate Governance and Accountability Review, vol. 1, No. 1. Retrieved from http://law.emory.edu/ecgar/content/volume-1/issue-1/essays/regulatory-capture.html.

Hofmann, S., and Ogonek, N., 2018. Different But Still The Same? How Public And Private Sector Organisations Deal with New Digital Competences. The Electronic Journal of e-Government, 16(2), pp. 127-135

IGI Global. (2020). What is Digital Illiteracy - IGI Global. Igi-global.com. Retrieved 18 March 2020, from https://www.igi-global.com/dictionary/strategic-crowdsourcing-as-an-emerging-form-of-global-entrepreneurship/51027.

Jakobsen, D., Jensen, M., & Tassy, A. (2018). IT-anvendelse i befolkningen. Danmarks Statistik.

Jensen, C., & Svendsen, G. (2011). Giving money to strangers: European welfare states and social trust. International Journal of Social Welfare, 20(1), 3-9.

Jobnet. (2020). Om Jobnet. Retrieved 28 April 2020, from https://info.jobnet.dk/om-jobnet

Kaelin, M. (2018). GDPR: A cheat sheet. Retrieved 28 November 2019, from https://www.techrepublic.com/article/the-eu-general-data-protection-regulation-gdpr-the-smart-personsguide/

Karaj, A., Macbeth, S., Berson, R., & Pujol, J. (2019). WhoTracks.Me: Shedding light on the opaque world of online tracking.

Kennedy, H., & Engebretsen, M. (2020). Data Visualisation in Society (1st ed.). Amsterdam University.

Kim, Y., Oh, T., & Kim, J. (2015). Analyzing User Awareness of Privacy Data Leak in Mobile Applications. Mobile Information Systems, 2015, 1-12. https://doi.org/10.1155/2015/369489

Kommunen. (2013). Kommunale indkøbsfællesskaber. Retrieved from https://www.kommunen.dk/kronik/kommunale-indkoebsfaellesskaber

Krishnamurthy, B. & Wills, C. (2006). Cat and mouse. Proceedings of The 15Th International Conference On World Wide Web - WWW. '06, 337-349. https://doi.org/10.1145/1135777.1135829

Krishnamurthy, B. & Wills, C.. (2010). On the Leakage of Personally Identifiable Information Via Online Social Networks. Computer Communication Review. 40. 112-117.

Kristensen, S. (2016). Advokaten 1 Når der kommer strøm til. Advokatsamfundet.dk. Retrieved 17 March 2020, from

https://www.advokatsamfundet.dk/Service/Publikationer/Tidligere%20artikler/2016/Advokaten%201/tema%20digitalisering.aspx.

Lauritsen, P. (2011). Big Brother 2.0: Danmark som overvågningssamfund. (1. udg.) København: Informations forlag.

Lehrer, C., Constantiou, I., & Hess, T. (2011). A cognitive processes analysis of individuals' use of location-based services. ECIS 2011 Proceedings. Paper 244.

Libert, T. (2014). Health Privacy Online: Patients at Risk. Data And Discrimination: Collected Essays, 11-15.

Libert, T. (2015). Exposing the Hidden Web: An Analysis of Third-Party HTTP Requests on One Million Websites. International Journal Of Communication.

Libert, T. and Nielsen, R. (2018). Third-Party Web Content on EU News Sites: Potential Challenges and Paths to Privacy Improvement. Factsheet. Reuters Institute for the Study of Journalism, University of Oxford. Retieved rom https://timlibert.me/pdf/Libert_Nielsen-2018-Third_Party_Content_EU_News_GDPR.pdf

Liebetrau, T. (2017). Digitalisering udfordrer statens sikkerhedsmonopol. Information. Retrieved from https://www.information.dk/debat/2017/10/digitalisering-udfordrer-statens-sikkerhedsmonopol

LinkedIn. (2019). Cookie Table. Linkedin.com. Retrieved 28 April 2020, from https://www.linkedin.com/legal/l/cookie-table.

LinkedIn. (2020). Cookie Policy. Retrieved 28 April 2020, from https://www.linkedin.com/legal/cookie-policy

Lyon, D. (2001). Surveillance after September 11. SAGE Journals. Retrieved 30 April 2020, from https://journals.sagepub.com/doi/abs/10.5153/sro.643.

Lyon, D. (2018). Surveillance capitalism, surveillance culture and data politics.

Margerie, E. (2018). Digital Illiteracy: OECD study highlights the incompetence of users. Marketing & Innovation. Retrieved 18 March 2020, from https://visionarymarketing.com/2018/10/digital-illiteracy-oecd-study-highlights-the-incompetence-of-users/.

Markham, A. & Buchanan, E. (2012). Ethical Decision-Making and Internet Research Recommendations from the AoIR Ethics Working Committee (Version 2.0). Retrieved from https://aoir.org/reports/ethics2.pdf

Marr, B. (2019). Chinese Social Credit Score: Utopian Big Data Bliss Or Black Mirror On Steroids. Forbes. Retrieved 9 March 2020, from https://www.forbes.com/sites/bernardmarr/2019/01/21/chinese-social-credit-score-utopian-big-data-bliss-or-black-mirror-on-steroids/#22e975d348b8.

Mathjax. (2020). MathJax. MathJax.org. Retrieved 28 April 2020, from https://www.mathjax.org/.

Mayer-Schönberger, V. & Cukier, K. (2013) Big Data: A Revolution That Will Transform How We Live, Work, and Think. Houghton Mifflin Harcourt.

Myers, C. (2018). Big Data and IT Infrastructure: Analyzing Connections to Boost Enterprise Security. Datafloq.com. Retrieved 9 March 2020, from https://datafloq.com/read/big-data-IT-infrastructure-analyzing-connections/5448.

Microsoft. (2020). IoT Edge | Microsoft Azure. Azure.microsoft.com. Retrieved 28 April 2020, from https://azure.microsoft.com/en-us/services/iot-edge/#iotedge-security.

Mitchell, R., Agle, B., & Wood, D. (1997). Toward a Theory of Stakeholder Identification and Salience: Defining the Principle of Who and What Really Counts. The Academy Of Management Review, 22(4), 853. https://doi.org/10.2307/259247

Nakashima, E., & Warwick, J. (2013). For NSA chief, terrorist threat drives passion to 'collect it all'. The Washington Post. Retrieved 4 May 2020, from https://www.washingtonpost.com/world/national-security/for-nsa-chief-terrorist-threat-drives-passion-to-collect-it-all/2013/07/14/3d26ef80-ea49-11e2-a301-ea5a8116d211_story.html.

New Relic. (2020). What is bam.nr-data.net? New Relic Browser monitoring. Retrieved 28 April 2020, from https://discuss.newrelic.com/t/relic-solution-what-is-bam-nr-data-net-new-relic-browser-monitoring/42055

Nordvpn. (2020). NordVPN | What is a VPN?. NordVPN. Retrieved 25 April 2020, from https://nordvpn.com/da/what-is-a-vpn/.

Nurse, J.R., & Buckley, O. (2017). Behind the scenes: a cross-country study into third-party website referencing and the online advertising ecosystem. Human-centric Computing and Information Sciences, 7, 1-21.

OAO. (2020). Det mener OAO om digitalisering. Offentligt Ansattes Organisationer. Retrieved 12 May 2020, from https://www.oao.dk/digitalisering/det-mener-oao-om-digitalisering/.

OECD. (2019). Social Expenditure Database (SOCX) - OECD. Oecd.org. Retrieved 8 May 2020, from http://www.oecd.org/social/expenditure.htm.

Omobowale, E., Kuziw, M., Naylor, M., Daar, A., & Singer, P. (2010). Addressing conflicts of interest in Public Private Partnerships. BMC International Health and Human Rights, 10(1).

Ontame.io. (2020). General Terms and Conditions. Retrieved 28 April 2020, from https://www.ontame.io/gtc18

Parkins, D. (2017). The world's most valuable resource is no longer oil, but data. Retrieved 26 February 2020, from https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

Persz, L. (2020). Stine Bosse: Vi må ikke blive så bange, at vi ikke tør tale om overvågningens bagsider. Politiken. Retrieved 4 May 2020, from https://politiken.dk/udland/art7750194/Vi-må-ikke-blive-så-bange-at-vi-ikke-tør-tale-om-overvågningens-bagsider.

Pew Research. (2016). Privacy and Information Sharing. Pew Research Center. Retrieved from https://www.pewresearch.org/about/our-mission/

Plesner, U. & Husted, E. (2019). Digital organizing: Revisiting central themes in organization studies. Basingstoke: Palgrave Macmillan.

Raymond, M., & DeNardis, L. (2015). Multistakeholderism: Anatomy of an inchoate global institution. International Theory, 7(3), 572-616. doi:10.1017/S1752971915000081

RawGit. (2018). RawGit. Rawgit.com. Retrieved from https://rawgit.com/.

Re, R. (2016). Imagining Perfect Surveillance. Papers.ssrn.com. Retrieved 1 May 2020, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2857234.

Regeringen. (2016). Redegørelse om Danmarks Digitale Vækst. Erhvervs- og Vækstministeriet. Retrieved from https://www.regeringen.dk/media/2069/redegorelse-om-danmarks-digitale-vaekst.pdf

Rey-Moreno, M., Felício, J., Medina-Molina, C., & Rufín, R. (2018). Facilitator and inhibitor factors: Adopting e-government in a dual model. Journal of Business Research, 88, 542-549.

Rucker, P., & Farnam, T. (2011). In Campaign 2012, Web sites are the new real estate. washingtonpost.com. Retrieved from https://www.washingtonpost.com/politics/in-campaign-2012-web-sites-are-the-new-real-estate/2011/10/17/gIQACSpssL_story.html.

Sanchez-Rola, I., Ugarte-Pedrero, X., Santos, I., & Bringas, P. (2017). The web is watching you: A comprehensive review of web-tracking techniques and countermeasures. Logic Journal of the IGPL, 25(1), 18-29.

Saunders, M., Lewis, P., & Thornhill, A. (2012). Research methods for business students (6th ed.). London: Pearson Education Limited.

Sawers, P. (2019). Realeyes Raises $12.4 Million to Help Brands Detect Emotion Using AI on Facial Expressions. VentureBeat. Retrieved from https://venturebeat.com/2019/06/06/realeyes-raises-12-4-million-to-help-brands-detect-emotion-using-ai-o n-facial-expressions/.

Scholl, H. (1970). Applying Stakeholder Theory to E-government. Towards The E-Society, 735-747. https://doi.org/10.1007/0-306-47009-8_54

Sharing Knowledge. (2020). Remove Pulseadnetwork.com, Ransom.Haknata.S1240226, NRnR.exe – Sharing knowledge. Sharingknowledge.world.edu. Retrieved 28 April 2020, from https://sharingknowledge.world.edu/remove-pulseadnetwork-com/.

Shih, C., Chen, F., Cheng, S., & Kao, D. (2019). Using Google Maps to Track Down Suspects in a Criminal Investigation. Procedia Computer Science, 159, 1900-1906. https://doi.org/10.1016/j.procs.2019.09.362

Siteimprove. (2020a). Siteimprove: Alt-i-en Website Management Software. Siteimprove. Retrieved 11 May 2020, from https://siteimprove.com/da-dk/.

Siteimprove. (2020b). Cookie Notice. Siteimprove. Retrieved 28 April 2020, from https://siteimprove.com/en/privacy/cookie-notice/.

SoundCloud. (2018). Cookies Policy. Siteimprove Retrieved 28 April 2020, from https://soundcloud.com/pages/cookies/03-2018

Stahl, L. (2018). Aleksandr Kogan: The link between Cambridge Analytica and Facebook. cbsnews.com. Retrieved from https://www.cbsnews.com/news/aleksandr-kogan-the-link-betweencambridge- analytica-and-facebook/.

Statsministeriet. (2017). Statsministeren nedsætter Disruptionrådet - Partnerskab for Danmarks fremtid. Statsministeriet. Retrieved from http://www.stm.dk/_p_14514.html

Surfing Waves. (2020a). Surfing Waves - Online Resource for Surf Addicts!. Surfing-waves.com. Retrieved 28 April 2020, from https://surfing-waves.com/.

Surfing Waves. (2020b). FREE Feed Widget. Fully Customisable and Easy To Use. Up to 5 feeds. Surfing-waves.com. Retrieved 28 April 2020, from https://surfing-waves.com/feed.htm.

Sweeney, L. (2013). Discrimination in online ad delivery. Communications Of The ACM, 56(5), 44-54. https://doi.org/10.1145/2447976.2447990

Thakuriah, P., Tilahun, N., & Zellner, M. (2017). Seeing Cities Through Big Data (1st ed.). Springer: Urban Big Data Centre.

Trzaskowski, J., & Sørensen, M. G. (2019). GDPR Compliance. København: Ex Tuto Publishing A/S.

UN. (1948). Universal Declaration of Human Rights.

UN. (2019). Data Economy: Radical transformation or dystopia?. The United Nations. Retrieved from https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/publication/FTQ_1_Jan_2019.pdf

Uzialko, A. (2018). How and Why Businesses Collect Consumer Data. Retrieved 28 November 2019, from https://www.businessnewsdaily.com/10625-businesses-collecting-data.html

Verdouw, C. N., Bondt, N., Schmeitz, H., & Zwinkels, H. (2014). Towards a Smarter Greenport: Public-Private Partnership to Boost Digital Standardisation and Innovation in the Dutch Horticulture. International Journal on Food System Dynamics, 5(1), 44-52.

VideoTool. (2020). VideoTool. Videotool.dk. Retrieved 28 April 2020, from https://www.videotool.dk/home.

Vrangbæk, K. (2009). Public Sector Values in Denmark: A Survey Analysis. International Journal of Public Administration: Public Values and Public Management, 32(6), 508-535.

Whitman, J. (2018). The currency of the modern world: Your attention. Retrieved 28 November 2019, from https://medium.com/@whitmaan/the-currency-of-the-modern-world-your-attention-25dfa724622a

WhoisGuard. (2020). Protect your privacy using WhoisGuard. Whoisguard.com. Retrieved 28 April 2020, from http://www.whoisguard.com/index.asp.

Wong, P. (2019). Everything a Data Scientist Should Know About Data Management*. Medium. Retrieved 11 March 2020, from https://towardsdatascience.com/everything-a-data-scientist-should-know-about-data-management-6877788c6a42.

Yang, S. O. (2016). Returning to the philosophical roots of sociomateriality: How M. Heidegger and M. McLuhan questioned information communication technology. ACM SIGMIS Database: The DATABASE for Advances in Information Systems, 47(4), 93–105. https://doi.org/10.1145/3025099.3025109

Yin, R. K. (2013). Case study research: Design and methods (5th ed.). Newbury Park, CA: Sage Publications Inc.

Yu, S., & Guo, S. (2016). Big data concepts, theories, and applications. Springer.

Zeller, F., 2017. Analyzing Social Media Data and Other Data Sources: A Methodological Overview. In: S. Luke and Q. Anabel, ed., The Sage Handbook of Social Media Research Methods. SAGE Publications Ltd, pp. 388-404.

Zuboff, S. (2015). Big other: Surveillance Capitalism and the Prospects of an Information Civilization. Journal of Information Technology, 30(1), 75–89. https://doi.org/10.1057/jit.2015.5

Zuboff, S. (2019). The age of surveillance capitalism: The fight for the future at the new frontier of power. London: Profile Books.

Appendix

Appendix A - SQL query for retrieving results for elements

```
SELECT *
FROM element
left join domain
on element.domain_id = domain.id
left join domain_owner
on domain.domain_owner_id = domain_owner.id
left join page
on element.page_id = page.id
where is_3p = 1
and domain not like '%naevneneshus.dk%'
and domain not like '%stil.dk%'
and domain not like '%lejre.dk%'
and domain not like '%vive.dk%'
and domain not like '%regionh.dk%'
and domain not like '%forsvaret.dk%'
and domain not like '%domstol.dk%'
and domain not like '%laegemiddelstyrelsen.dk%'
and domain not like '%esbjerg.dk%'
and domain not like '%ddsks.dk%'
and domain not like '%tbst.dk%'
and domain not like '%soroeakademi.dk%'
and domain not like '%svs.dk%'
and domain not like '%simb.dk%'
and domain not like '%sim.dk%'
and domain not like '%regionsyddanmark.dk%'
and domain not like '%oim.dk%'
and domain not like '%oes.dk%'
and domain not like '%kultunaut.dk%'
and domain not like '%kefm.dk%'
and domain not like '%justitsministeriet.dk%'
and domain not like '%hvidovrehospital.dk%'
and domain not like '%herlevhospital.dk%'
and domain not like '%hedw.dk%'
and domain not like '%fmn.dk%'
and domain not like '%ejendomsstyrelsen.dk%'
and domain not like '%detgroennemuseum.dk%'
and domain not like '%bronderslev.dk%'
and domain not like '%bispebjerghospital.dk%'
and domain not like '%at.dk%'
```

Appendix B - SQL query for retrieving results for cookies

```
select *
from cookie
left join page
on cookie.page_id = page.id
```

Appendix C – Full WebXray SQLlite file output

- https://github.com/TeisLebeck/TPT-Danish-public-web-pages?fbclid=IwAR3M82b1FzuaDot0JnUewCZhHxkubSqfaNcsQIgDsA13OSKDaK-kakvi348