NTNU
Norwegian University of
Science and Technology

# Privacy and Social Media: Do Users Really Care?


Hannah Ersdal
Sølvi Svendby Skjærstad



Master of Science in Communication Technology
Submission date: June 2016
Supervisor: Maria Bartnes, ITEM
Co-supervisor: Lillian Røstad, ITEM



Norwegian University of Science and Technology
Department of Telematics

**Title:** Privacy and Social Media: Do Users Really Care?

**Student:** Hannah Ersdal and Sølvi Svendby Skjærstad

**Problem description:**

Privacy concerns the individual's right to control his/her own personal information. The collection of personal information shall have a clear purpose, and there are a number of other principles as well, for access to, and deletion of, personal information, among others. Social media provide platforms for us to share personal information as part of our social interaction with other people. Most of these online services and platforms are also freely available. However, our personal information tends to be collected and used commercially (not always in the most transparent way), and this is the price we pay for free access.

In this master thesis, we will be focusing on the modern use of social media. Our aim will be to look into at what degree social media users care about their privacy issues when sharing personal information via social media platforms. In the means of investigating to what extent users share information with other users, and to what extent they are aware of the information social media owners collect and (also) sell to third-party companies.

In addition, our intention is to gain insight into social media users' awareness of their personal information being used for commercial use. To what degree would such information affect their willingness to share personal information?

Lastly, our aim is to investigate if a person's opinion of privacy affects user behavior. Do the users really read and consider the content of privacy policies and do they actively take control of the privacy settings?

**Responsible professor:** Maria Bartnes, ITEM

**Supervisor:** Lillian Røstad, ITEM

## Abstract

Over the last decade, social media networks have experienced explosive growth. Social media has become a common form of communication for most people, and the average person spends more and more time in front of the computer. We are exposing ourselves online, and consequently leaving more personal information on the Internet than ever before. Resulting in personalization and individualism being the drivers of the networks, and this has been made possible by huge amounts of data.

The thesis studies to what degree users care about privacy on social media platforms. To investigate at what extent users share information and are aware of information being shared with third-party companies, we constructed the following research questions;

1. Do social media networks protect the personal information of their users in the same fashion or are there any differences?
2. Other than the social media network itself, who else collects information about its users and how is the information spread between parties?
3. What do social media users know in the terms of how and how much information is being spread? Do they care?

Through a documentation analysis, the thesis has examined the documentations provided by different social media networks. The thesis has analysed different third-party companies present on various websites. A mapping of these finding was done to illustrate the large web they conclude. Through conducting a user survey, the thesis gained insight into Norwegian social media users' habits and their knowledge concerning the discussed topics.

An evaluation of our findings ultimately leads to the conclusion of the privacy paradox holding true for the users involved in our study. We found that users claim to care about privacy online and that many have knowledge concerning the aspects analysed. However, they do not read the documentation and still utilise the services provided without having a clear understanding of how the technologies work on the Internet.

## Sammendrag

I løpet av det siste tiåret har sosiale medier hatt en eksplosiv vekst. Sosiale medier har blitt en vanlig kommunikasjonskanal for folk flest og den gjennomsnittlige personen bruker mer og mer tid foran datamaskinen. Vi eksponerer oss selv på nettet og legger igjen mer personlig informasjon enn noen gang før. Dette fører til personalisering og individualisme som drivere av sosiale medier og dette har blitt muliggjort ved store datamengder.

Denne oppgaven studerer i hvilken grad brukere bryr seg om personvern på sosiale medier. For å undersøke hvor mye informasjon brukere selv deler og hvor mye de vet om deling av informasjon til tredjepartsbedrifter, utformet vi følgende problemstillinger:

1. Beskytter sosiale medier brukernes personlige informasjon på samme måte eller er det forskjeller?
2. I tillegg til de sosiale mediene, hvilke andre bedrifter samler informasjon om brukerne deres og hvordan blir denne informasjonen delt mellom de?
3. Hva og hvor mye vet sosiale mediebrukere om deling av informasjon? Bryr de seg?

Gjennom en dokumentasjonsanalyse har oppgaven undersøkt dokumentasjon gitt av forskjellige sosiale medier. Oppgaven har også analysert noen av tredjepartsbedriftene som er tilstede på forskjellige nettsteder. En kartlegging ble gjort av disse funnene for å illustrere det store nettverket de utgjør. Ved å gjennomføre en brukerundersøkelse gir oppgaven innblikk i norske sosiale mediebrukeres vaner og deres kunnskap om de diskuterte temaene.

En evaluering av våre resultater har ført til en konklusjon som tilsier at personvernparadokset stemmer for brukerne involvert i denne studien. De sier at de bryr seg om personvern på nettet og mange har kunnskap om de ulike analyserte aspektene. På den andre siden leser de ikke dokumentasjon samtidig som de fortsetter å bruke tjenestene som tilbys på nettet uten å ha en tydelig forståelse av hvordan teknologiene fungerer.

# Preface

This master thesis concludes our Master of Science degree in Communication Technology at the Norwegian University of Science and Technology (NTNU). Both authors specialise in the field of Digital Economics at the Department of Telematics (ITEM) at the Faculty of Information Technology, Mathematics and Electrical Engineering (IME).

Firstly, we would like to thank our responsible professor Maria Bartnes, at the Department of Telematics (ITEM), for her contribution and guidance throughout the semester.

We would also like to thank our supervisor Lillian Røstad, at the Department of Telematics (ITEM), for valuable input in the toughest of times.

A huge thanks goes out to all who participated in our user survey for your vital contribution.

Lastly, we would like to thank the wonderful girls at Casa Rosa, Torres, and Mr. Avocado for inspiration this concluding semester at NTNU.

Hannah Ersdal and Sølvi Svendby Skjærstad
Trondheim, June 2016

# Contents

# List of Figures

# List of Tables

# List of Acronyms

**EEA** European Economic Area.

**EFF** Electronic Frontier Foundation.

**EU** European Union.

**GPS** Global Positioning System.

**HTML** HyperText Markup Language.

**HTTP** Hypertext Transfer Protocol.

**IME** Faculty of Information Technology, Mathematics and Electrical Engineering.

**IoT** Internet of Things.

**IP** Internet Protocol.

**ITEM** Department of Telematics.

**NSA** National Security Agency.

**NSD** Norwegian Centre for Research Data.

**NTNU** Norwegian University of Science and Technology.

**PDA** Personal Data Act.

**PDR** Personal Data Regulations.

**PII** Personally Identifiable Information.

**SNS** Social Networking Site.

**URL** Uniform Resource Locator.

**US** United States.

# Chapter 1
# Introduction

Over the last decade, social media networks have experienced explosive growth. Social media has become a common form of communication for many people, and the average person spends more and more time in front of the computer. People are using the Internet to do everyday things, i.e., shopping, reading news articles, watching TV-series and movies, talking to people, and listening to music. All in all, by this we are exposing ourselves online, and the consequences are that we are leaving more information about ourselves on the Internet than ever before.

Usage of social media has also gone through a change during the last years. At the outset, it was a channel for sharing interests and ideas. Now, it has become a huge platform with endless opportunities for both individuals and businesses. The basics still apply, but social media networks now offer companies opportunities to connect with individual customers. Resulting in personalization and individualism being the drivers of the networks [1], and this has been made possible by huge amounts of data.

Everything is, or can be made, available on the internet. An example of this was from 2013 when Edward Snowden leaked classified information from the National Security Agency (NSA) [2]. The uproar around this event was for many about how the United States (US) could possess so much information about people and various nations. There were also, thankfully, several reactions to how one man could get access to this much information.

Moreover, what has been done cannot be undone. When something has been posted on the Internet, it is there forever. This applies as much to the information the average user provides to social media networks as it does for the documents Snowden released in 2013. Somewhere there will always be a backup, and someone will always be able to trace back to, or restore, the data.

Privacy has therefore never been more important. Are we safe online? People claim to care about privacy, and that they are concerned with this topic when

using the Internet. However, how much does the average user know about what is happening to the information they publish? The Norwegian Data Protection Authority frequently releases reports and recently provided a report called The Great Data Race. The topic was how commercial utilisation of personal data challenges privacy. At what extent do these reports reach the public? More importantly, would the average user read and understand the contents?

Social media networks are, as mentioned, also using personal information for commercial utilisation. After searching for new shoes online, a user will experience that the same shoes show up on their Facebook newsfeed [3]. Meaning, Facebook does not only collect and store the information users provide directly to the service but also tracks users across the Internet for advertising purposes. What is the extent of social media networks tracking online? How much information do they collect and how much is shared with other parties?

This thesis will look into social media networks, how they protect user information and how information about users is shared online. Combining this information with the knowledge users hold regarding these topics will let us investigate whether the users care about privacy in social media.

## 1.1 Objectives

The goal of this master thesis is to study to what degree users care about privacy on social media platforms. In the means of investigating to what extent users share information and are aware of information being shared with third-party companies. This leads to the following research questions, which, ultimately, define our objectives:

1. *Do social media networks protect the personal information of their users in the same fashion or are there any differences?*
2. *Other than the social media network itself, who else collects information about its users and how is the information spread between parties?*
3. *What do social media users know in the terms of how and how much information is being spread? Do they care?*

## 1.2 Scope and Limitations

Social media is a broad term, and various social media networks have emerged through the years. This master thesis will concentrate on four of the most well-known platforms, namely, Facebook, LinkedIn, Google+, and Twitter. We are focusing on these as they are the largest in Norway, with the highest rate of daily activity[4].

We limit the scope of our study to social media networks that are accessible by web applications, consequently emitting services such as Instagram and Snapchat. These are among the biggest social media networks in Norway considering the number of users [4] but are mainly used as mobile applications.

Concerning information sharing between third-parties, we limit the tests to include 22 websites. More than this would have been too extensive for the thesis as this part was executed to get an initial understanding of sharing of information online. For this reason, we also limit the number of testing tools to two user-friendly browser extensions.

To gain insight into social media users' knowledge and awareness, we conducted a user survey. The aim was to reach out to members of the chosen social media networks of all ages. Additionally, we wanted to limit the scope to Norwegian users and, therefore, provided the user survey in Norwegian.

## 1.3 Contribution

The contribution of this thesis is the evaluation of privacy policies on social media networks and third-party trackers on popular websites. The main contribution, however, is the investigation of an average Norwegian user's understanding surrounding the topics. Combining these two factors, we hope that the thesis will be of value to both social media users and companies.

## 1.4 Outline

The thesis is structured into eight chapters, and the outline is as follows:

- Chapter 1, Introduction: contains the motivation and objectives for the thesis. The chapter also includes scope and limitations, and contribution.

- Chapter 2, Background: presents the necessary background material for the thesis. Insight is given into social networking, online tracking mechanisms, internet economy, and privacy.

- Chapter 3, Methodology: includes a description of the research methods used and challenges that may arise.
- Chapter 4, Comparison of Privacy Policies: evaluates and compares privacy policies of four social media sites.
- Chapter 5, Mapping of Third-Party Trackers: includes the testing of 22 websites and mapping of trackers on these sites. The chapter also includes insight into how we executed the tests, the results, and briefly examines the dominating tracking companies discovered.
- Chapter 6, Test of User Knowledge: presents the findings from the user survey.
- Chapter 7, Discussion: summarises and discusses the results found in the thesis. A brief discussion of limitations experienced is also presented.
- Chapter 8, Concluding Remarks and Further Work: concludes the thesis and proposes further work.

# Chapter 2
# Background

The aim of this chapter is to provide the reader with insight regarding information collection, tracking technologies used or present online, the state of the Internet economy, and the privacy considerations concerning these topics.

The following section includes an introduction to different social media sites and the types of data they collect from users. Further, we take a closer look at some standard online tracking methods and how to avoid them, before moving on to a presentation of the status of Internet economy today. Lastly, we discuss the privacy term, and issues and legal aspects that arise with it.

## 2.1 Social Networking

Social platforms, or social media sites, have become increasingly more popular over the years. Every day millions of people use sites such as Facebook, Instagram, Snapchat, and LinkedIn, among others, to communicate with friends, family and co-workers. As of April 2016, Facebook registered more than 1.6 billion monthly users [5]. To put that number in perspective, this is more than the current population of China (1.38 billion) and over three hundred times as many as the people of Norway (5.084 million) [6]. Undoubtedly, social networking is playing a significant role in our daily lives.

Social media sites differ from each other in various ways, regarding design, purpose, and functionality. Common for most sites, however, are that they allow users to create personal profiles, publish content, and connect to other users. Users often have the opportunity of creating groups where people with similar interests can join and interact with each other by, for example, sharing information or create events.

Classifying the different social media sites makes it easier to both separate and understand the variations. Following is a much-used classification [7].

- **Networks** - Social networks include services that allow users to connect with other people of similar interests and background. Such networks can be professional (e.g., LinkedIn) or social (e.g., Facebook). The websites usually consist of personal profiles and different ways of communicating and sharing content with others.

- **News** - Social news sites allow users to share various news items or links to outside articles and also vote on the different links and items. The "core social aspect" is thus the voting as the elements with the highest number of votes are most prominently displayed [7]. Reddit is an example of a social news site.

- **Microblogging** - A microblog is a type of blog that lets users publish updates to anyone subscribed to receive them. An example of this kind of social media is Twitter. The updates are usually short and limited to a particular word count.

- **Media sharing** - Media sharing websites include services that allow users to share different types of media (i.e., videos and images) with other users. These sites usually offer social features, such as creating profiles, commenting on posted media, or send messages. Examples in this category are Flickr and YouTube.

- **Bookmarking sites** - These sites allow users to bookmark, i.e., save and organise websites they enjoy. A popular feature lets users "tag" the sites they wish to save, making them easy to search for or share. StumbleUpon is an example of a bookmarking website.

- **Forums and blog comments** - Forums are online platforms that allow users to hold conversations by posting and responding to messages. A blog comment site is a bit more focused than forums as the comments are often centred around the subject of the blog post.

It is important to note, however, that there are no strict boundaries between the different categories. Social media sites may implement features of various categories and thereby overlap with regards to definitions. Some examples of this include Facebook and Twitter. Twitter is often considered a microblog but offers features which make it definable as a social network as well. Similar, we find Facebook whose "status update" feature resembles the aspect of microblogging even though it usually is defined as a social network.

As of now, a social media network will be referred to as a Social Networking Site (SNS) in this thesis.

### 2.1.1 Information Collected by Social Networking Sites

Information about Internet users has become a commodity [8]. Everything we do online is being tracked and monitored by different actors. We will elaborate on who these actors are in Section 2.3 and further analyse the third-parties who track us on various websites in Chapter 5.

It is not uncommon for people to have more than one SNS account. People are often not aware how much the networks know about them or how much information they are providing the services. To get some understanding of this, we are including a list presenting an overview of data types SNSs may collect about their users. There are many different taxonomies on the Internet regarding information collected by SNSs. Consequently, the list provided below is based on the 2010 version by Bruce Schneier [9] and the revised 2014 version by Richtammer et al. [10]. This classification will, consequently, be used throughout this thesis.

- **Service Data** consists of information users provide to a SNS about themselves in order use or enhance the use of the service. We define two types of Service Data;
  - **Mandatory Service Data** is the minimal amount of information required from users for them to be able to use the service, and often includes information such as legal name, age, gender, and similar. Included in Mandatory Service Data is *Login Data*, meaning credentials needed to sign into the service. This typically includes username, password, email address, phone number, and similar.
  - **Voluntary Service Data**, sometimes referred to as *Extended Profile Information*, is any additional information the users choose to give about themselves. This can be interests, workplace information, and so on.
- **Disclosed Data** refers to content the users themselves post or share on their profiles or pages. This includes photos, status updates, videos, text posts, and so on.
- **Entrusted Data** is content users themselves post on other people's pages or profiles, meaning that the user is not in control of the published content once it is posted because it is part of the other user's account. Entrusted Data includes the same type of content as Disclosed Data.
- **Incidental Data** refers to the information other people post or share about the user. In addition to not having control of the information once it is posted, the respective user did not create it in the first place, i.e., is not the sole owner. We separate between two types of Incidental Data.

    - **Contextual Data** includes the same type of content as Disclosed Data.
    - **Private Communication Data** is content collected from private messages, video chats, InMail (a LinkedIn solution), and similar.

- **Behavioural Data** refers to information about users' behaviour and navigation on SNSs, and information collected from a user's interaction with third-party applications. This includes information such as pages visited, news articles accessed, games played, topics written about, and similar.

- **Connection Data**, also known as *Log Data*, is technical information generated by the platforms users use to access the services. We define two types of connection data;

    - **Device Data** is information regarding devices and technologies used to access services and includes information such as browser type, operating system, mobile device(s), Internet Protocol (IP) address, and so on.
    - **Location Data** is derived from the user's IP address or by using Global Positioning System (GPS).

SNSs may collect one or several of the data types described. When they do, information is associated with the users' respective accounts and used for various reasons. We will come back to this in Chapter 4. Consequently, SNSs have the potential of processing massive amounts of information about its users. This has led to an increasing interest in information trading and the value of user information [1]. We will come back to this in Section 2.3.

## 2.2 Online Tracking

Many different mechanisms provide tracking of online users. SNSs often offer their users their services for free. This is usually a result of services exploiting their users in other ways. Personalised content and advertising require information on the individual user, and users pay for services with personal information [1].

According to the Norwegian Data Protection Authority, this development is driven by trends such as Internet of Things (IoT) and wearable technology, e.g., smart watch, smartphones. This evolution opens up for new possibilities for information collection as these gadgets are becoming a part of our everyday life. Geographical location and health information, such as heart rate or activity monitoring, may be collected using various new tracking mechanisms. Online, however, we have more "traditional" tracking mechanisms. These are built around the use of web browsers on computers, and they are the ones investigated in this thesis.

Online companies can track users over multiple websites for various reasons. Tracking technologies can be used for personalisation, meaning that a site remembers a user's login credentials. They can also be used to, for example, remember items placed in "shopping carts" when shopping online, called session management. Additionally, tracking technologies can be used to store information regarding a user's web browsing habits [11].

The following sections will present some of the most frequently used tracking technologies and shortly explain what Internet users can do to limit information collection about them.

### 2.2.1 HTTP Cookies

Hypertext Transfer Protocol (HTTP) Cookies is the most popular technology when it comes to tracking users online [1]. Using cookies entails that when a user visits a website, a piece of code is stored in the user's web browser, ensuring that the web browser, or device, is recognised if the user returns to the site at a later time.

When reading cookie policies of popular SNSs, two types of cookies are repeatedly mentioned, namely persistent and session cookies. Session cookies, also known as temporary cookies, are only active for one session, meaning that they are deleted, i.e., expire, when the user closes the web browser. In contrast to persistent cookies which are stored on the user's browser until it expires at a specific date or after a length of time [11].

Advertisers can use persistent cookies to collect information about a user's browsing habits and because of this, persistent cookies are often referred to as tracking cookies. Tracking cookies are frequently discussed concerning privacy issues, and this especially involves third-party tracking cookies. Third-party cookies are cookies set out by someone other than the domain owner of the respective website, e.g., advertising companies or data brokers. These are often used to track users over a longer period to create user profiles, as mentioned in Section 2.3.1. These profiles also include information about the user's online behaviour. For countries within the European Union (EU), websites are obliged to inform its visitors of the presence of such cookies [11].

### 2.2.2 Social Widgets

Another popular technology is web widgets. Widgets are small pieces of code that are placed on websites to interact with, display content from or redirect users to other websites or applications. They are often referred to as self-contained code, meaning that they are small applications that open up doorways to much larger applications [12]. Typical widgets include dialogue boxes, pop-up windows, forms,

or buttons. They may provide search boxes for Google or any other search-based website, weather forecasts, games, or social media content.

Companies typically use widgets to enhance their websites [13]. News sites, for example, may place social sharing buttons on their site for the possibility to reach out to a larger audience.

Social widgets, or social plugins, are usually provided by SNSs themselves and collect information about user behaviour, as was the case with cookies. It is important to note that if a user has an account on a SNS, the SNS will collect information about all the websites the user visits that have included the respective social media widget [13]. We will come back to the different types of widgets available from some of the largest SNSs in Chapter 4.

### 2.2.3 Other Mechanisms

In addition to the technologies mentioned above, there are numerous other ways and variations for companies to track Internet users. Worth mentioning are digital fingerprinting, web beacons and HyperText Markup Language (HTML) Local Storage.

#### Digital Fingerprinting

A digital fingerprint is a term for the unique electronic "fingerprint" every device has when connected to the Internet [1]. It is composed of several elements, and can, therefore, provide detailed information about a user. Such information often includes IP address, browser type and software used, device information, and other settings such as language [8].

#### Web Beacons

Web beacons are transparent graphic images placed on websites either to collect information or place cookies. They are often used in combination with cookies to collect additional information. This information may include IP address, type of web browser used, and the time the user visited the website.

#### HTML Local Storage

HTML Local storage is very similar to persistent cookies but differ in the fact that the storage limit in local storage is a great deal larger, as they can store at least 5MB [14]. Cookies are sent with each HTTP request, hence the name HTTP cookies, and this can slow down the visited website. Contrary to HTML local storage where running time will not be affected in the same way as it is only delivered to the site when it is specifically requested.

### 2.2.4 Privacy Enhancing Mechanisms

Users have some options when it comes to limiting the information collected, and to getting an overview over companies and their reasons for tracking.

Do Not Track is a browser setting users can utilise signalling companies that they do not want their behaviour to be recorded and collected. The Do Not Track setting is a HTTP-header that sends out a signal of 0 or 1, depending on the user's wishes. Do Not Track is only a proposed header, meaning that there is no standard telling companies how to interpret the signals. Therefore, it is currently up to each company to decide what to do when they receive a Do Not Track signal. Reportedly, most websites have not changed their practices yet and will probably not consider it until a standard is in place [15].

Another option for users is to manually opt-out from being tracked by companies. Ironically, this is accomplished by installing a cookie, often referred to as an opt-out cookie, in their browser. Opt-out cookies prevent future cookies from being installed in the user's browser and are usually provided by the companies setting out tracking cookies in the first place [16].

Additionally, many browser extensions allow users to block or identify tracking companies. In Chapter 5, we discuss two of these and use them to map third-party trackers present on various websites.

Lastly, it is important to include that tracking and the use of cookies, plays an important role for website owners. Every website is dependent on knowledge about what visitors are doing on their site to be able to develop and provide both tailored and personalised content, as well as personalised experiences and advertising. Cookies and other tracking mechanisms are important parts of the Internet. Without them, online companies would not be able to know how to make their services more attractive for their visitors and websites would be a lot less interactive [17].

## 2.3 Internet Economy

Over the past six months, the Norwegian Data Protection Authority have published two reports on the current state and trends regarding privacy online and the expected trends of 2016 [8] [1]. By discussing the state of the Internet economy, with a description on tracking of Internet users and how automated ad trading works, they provide readers with valuable insight into today's practices.

Personal information online is, as mentioned previously, considered as a commodity. One of the drivers for this development is the increasing trend of Big Data analysis [8], e.g., data mining. Data mining is the practice of looking for correlations in and

organising large quantities of unstructured data. Another driver is that companies now can store more data than before due to the steady increase in storage capacity and cheaper computing power [8].

The advertising industry is taking advantage of this development, and automated advertising (ad) trading is now a common practice [1]. The following section will explain this process further, in addition to giving an overview of the different actors involved.

### 2.3.1 Automated Advertising Trading

As mentioned, the Norwegian Data Protection Authority's recent work emphasises the current practices of automated ad trading. Broadly speaking, the market consists of buyers of advertising space on one side and sellers, or vendors, of space on the other. It can be challenging making a clear separation between the actors because they sometimes cover several roles at the same time [1]. We will get back to this later in this section.

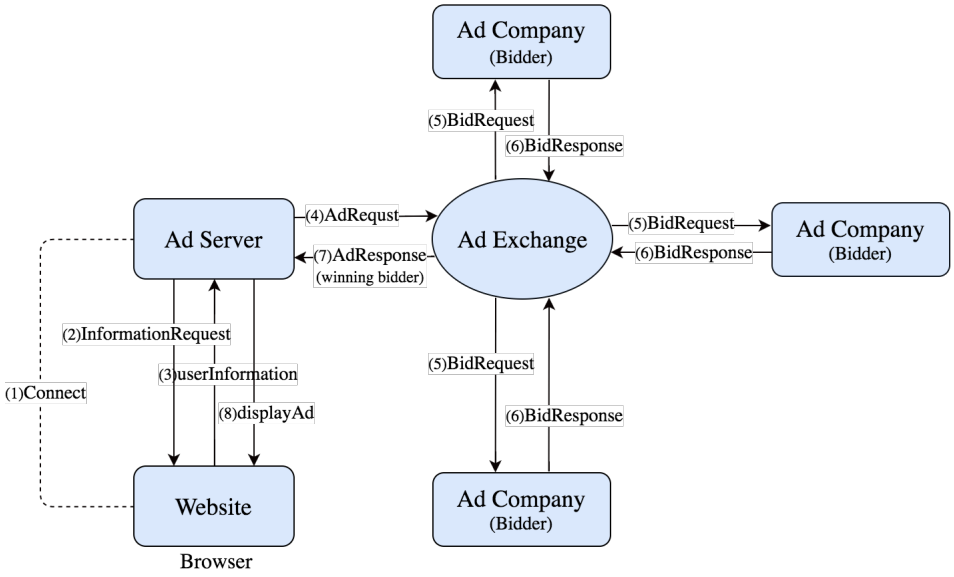


**Figure 2.1:** Flow Diagram for Automated Ad Trading

Before explaining the roles of the most prominent actors, let us examine what occurs "behind the scenes" when a user visits a website. The information flow is illustrated in Figure 2.1, and is based on the process description from the report by the Norwegian Data Protection Authority [8]. Note that all of this is happening

within milliseconds and occurs from when a user enters a website's Uniform Resource Locator (URL) until the website is fully loaded in the browser.

Firstly, a connection is established between the user's browser and an advertising server. The server informs the website owner to fill the initially empty advertising space with content. Then, the site sends a message, through the server, to an ad exchange, which invites ad space buyers to place a bid on the user in question. Registered advertising companies receive information about the user from the ad exchange, which may include the user's IP address, location, gender, assumed interests, income, in addition to the website the user is visiting [8].

Advertisers combine this information with information they may already have about the user. This results in an algorithm calculating whether a bid should be placed, and, if so, how high it should be. Finally, the advertising company with the highest bid wins the right to show the user an advertisement.

Figure 2.2 displays the value chain for automated ad trading. Here we have an overview of the various actors involved in the trading process and how they are positioned in correlation to each other [8]. The following sections include explanations to each actor and their role in turn.

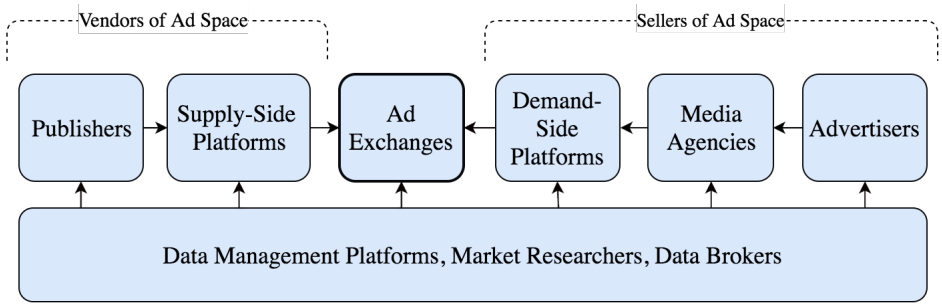


**Figure 2.2:** Value Chain for Automated Ad Trading

### Advertising Exchanges

*Ad exchanges* are located in the middle of the value chain. These are marketplaces for purchase and sale of advertising space, and they build on the same principles as stock exchanges. Ad exchanges serve as a neutral platform where advertisers can bid on users posted by publishers in real time, i.e., a platform for real-time bidding.

### Vendors of Advertising Space

To the left, we find vendors of ad space, i.e., *publishers* and *supply-side platforms*. Publishers make their living from selling ad space on their sites to advertisers and do this by exploiting supply-side platforms [8]. A supply-side platform is a software specially developed for this purpose, and they deliver information about users to the ad exchange [1].

### Buyers of Advertising Space

Purchasers of advertising space are located to the right, and include *advertisers*, *media agencies*, and *demand-side platforms*.

As we remember from Section 2.1.1, SNSs associate information about users with each user's respective account. Similarly, advertisers create profiles on each user [1]. These may help advertisers recognise which users are most likely to buy their products, and consequently, on which users to place higher bids.

Advertisers wanting to buy advertising space use a demand-side platform, similar to how publishers use supply-side platforms. Demand-side platforms are typically operated by media agencies or large companies such as Google or Yahoo [8]. An algorithm, developed in cooperation with the advertiser, determines whether or not the user is valuable for the advertiser.

### Data Management Platforms, Market Research, and Data Brokers

The last group of actors include companies that make a living off selling user profiles, data and market analysis to both publishers and advertisers. They make up the largest group of third parties present on websites. *Data Brokers* collect users' personal data, often by placing cookies in their browsers, and resell or share this information with others [1].

*Data management platforms* include companies that offer tools for both analysing data and purchasing ad space. Information from the demand-side platform may be sent to and combined with information from the data management platform and thereby used to develop ad-targeting algorithms.

Lastly, *market research* companies contribute by finding the target group for advertising and evaluate the effects of marketing campaigns. Typically, information is collected using web panels or telephone interviews [1].

The process and basics of advertising trading may look simple. The reality is, however, that there are hundreds of companies competing. Additionally, it may be difficult to provide a clear separation of the various actors. This is, as mentioned,

due to the fact an actor may cover more than one role in the trading process. For example, in some cases, Google may be a publisher in addition to providing tools for both the supply and demand side platform, e.g., Admeld and DoubleClick Bid Manager, respectively.

## 2.4 Privacy

In the previous sections, we have been introduced to how companies collect information about Internet users and types of data collected by SNSs. With the collection of personal information, the concern for online privacy arises. In the following sections, we will give insight into privacy and legal aspects that occur when sharing information on the Internet.

Privacy is a complicated concept and depends on the situation at hand. The English dictionary defines privacy in four ways [18]:

1. "The state of being apart from other people or concealed from their view; solitude; seclusion"
2. "The state of being free from unwanted or undue intrusion or disturbance in one's private life or affairs; freedom to be let alone"
3. "Freedom from damaging publicity, public scrutiny, secret surveillance, or unauthorized disclosure of one's personal data or information, as by a government, corporation, or individual"
4. "The state of being concealed; secrecy"

In this thesis, we are considering privacy on the Internet. This may be a different way of thinking about it, but the basics are the same. The concept still constitutes the protection of an individual's integrity. Privacy concerns the individual's right to control his/her personal information and is recognised as a fundamental human right.

### Privacy on the Internet

On the Internet, the privacy concept concentrates around the protection of a user's personal data. SNSs provide platforms for users to share personal information as part of their social interaction with other people. Usually, websites owners want to provide visitors with user-friendly and tailored experiences. To do so, they exploit information regarding the visitors' online behaviour.

All in all, a lot of Internet activity revolves around the collection of personal information. Companies that live off collecting user data need to present users with information about how and why they are doing so. Consequently, many websites

provide their visitors with privacy policies. This is done to give an understanding of and a clear purpose for their data collection.

A privacy policy states how a company collects information, what is collected, and how information is used. Firstly, privacy policies need to include what information they gather, whether it be Service or Behavioural Data. Another important part is how the information is collected and whether the websites leave data on the computer to gather the information, i.e., by use of cookies. Lastly, privacy policies need to include what the gathered information is used for and who else potentially receives it.

Larger companies, such as Facebook and Google, provide their users with numerous policies referring to different products and services. We will get back to this topic in Chapter 4.

Many online services and platforms are freely available for users. However, the users are most likely paying with personal data. It is important to understand, however, that there are differences in the level of personal information websites collect. The information does not necessarily need to be what is known as Personally Identifiable Information (PII), but it is safe to say that a lot of information about users is collected and used for commercial reasons all over the Internet.

### Personally Identifiable Information

So far in the thesis, we have referred to the term "personal information" several times. We, therefore, find it important to define what this entails, and the meaning of Personally Identifiable Information (PII).

The Norwegian Data Protection Authority defines personal information to include, but not limited to, name, address, telephone number, email address, IP address, vehicle registration plate number, and fingerprint [19]. The National Institute of Standards and Technology classifies this information PII, as well. By PII, we mean information that may directly, or indirectly, identify an individual by one or more factors specific to "physical, mental, economic, cultural, or social identity" [20]. Additionally, Behavioural Information, as described in 2.3, is considered to be personal information by the Norwegian Data Protection Authority.

### The Privacy Paradox

In today's society, we have what is called the privacy paradox, which involves the "relationship between individuals' intentions to disclose personal information and their actual personal disclosure behaviours" [21]. In simpler words, the paradox suggests that while Internet users claim to care about privacy, their behaviour says

otherwise. A great number of surveys conclude that people express to care about privacy [22], though the majority chooses convenience and connectivity over the alternative. It seems that people willingly offer privacy for the other goods the Internet provides.

However, do users have a choice when it comes to offering privacy? Website owners often go by the rule of users having to accept their terms if they want to continue using their site. It is easier just to agree to "I have read and agree to the terms of use" than to read them before accessing the service. This results in users having little knowledge of what information companies collect and how they use this information. In Chapter 6, we will come back to this topic by providing insight into Norwegian SNS users, and their knowledge regarding online privacy and information sharing.

### 2.4.1 Legal

When it comes to privacy online, there has evolved the need for specific laws for the protection of personal data. Norway currently follows two laws concerning this topic, in addition to adhering to European regulations. In the following section, we will briefly look into how an individual's personal data is protected online. We then move on to an introduction to how data is transferred between countries. Lastly, we will look into new regulations for the protection of personal data that are to take effect in Norway by 2018.

#### The Transfer of Personal Data to Other Countries

For Norwegian citizens, the transfer of personal data to other countries is protected by the Personal Data Act (PDA) ("Personopplysningsloven") and the Personal Data Regulations (PDR) ("Personopplysningsforskriften"). Companies wanting to transfer personal information to countries outside of Norway may only do so if they assure the adequate level of protection stated in the regulations.

The purpose of the PDA is to protect people from violation of their right to privacy through the means of processing personal information [23]. The PDA states that the transfer of personal data needs to happen with an adequate level of protection. EU/European Economic Area (EEA) countries are recognised as countries which maintain this level of protection, and transfer between these countries can, therefore, be done without any additional terms. This also applies to countries approved by the European Commission.

### Safe Harbor

Rules provided by the Data Protection Directive protects personal data in EU/EEA countries. As mentioned, these rules are implemented to Norwegian law through the PDA. The US, however, do not comply with the same regulations for the protection of personal data [24]. Because of this, the Data Protection Directive prohibits the transfer of personal data between the EU and the US unless consent is given or additional terms are met.

The purpose of the Safe Harbor Privacy Principles is to make sure companies in the US transfer personal data in a secure way [24]. The Principles are an agreement between the EU and the US. Businesses in the US following these principles are considered to have the adequate level of protection for the transfer of personal data between EU and US. This means that the Safe Harbor Privacy Principles also regulate how companies can transfer personal data from Norway to the US.

On October 6th, 2015, the European Court of Justice declared the Safe Harbor Privacy Principles as invalid [25]. Companies wanting to transfer personal data from the EU to the US now need to make use of other mechanisms. Existing standard contracts have to be used when considering data export. In all, there are three such contracts given by the European Commission. These have been the recommended legal basis when transferring data to countries without the adequate level of protection of personal data for several years [25].

### EU-US Privacy Shield

On February 2nd, 2016, the European Commission and the US agreed on a replacement for the EU-US Safe Harbor Privacy Principles [26]. The new framework for transatlantic flows will be known as the EU-US Privacy Shield. However, when the new principles will be finalised is yet to be declared.

### The Data Protection Regulation

The European Parliament have, as of April 14th, 2016, finalised and replaced the EU data protection directive from 1995. The goal of the regulation is to strengthen the trust for and provide a higher level of protection for individuals across the EU [27]. This will apply to companies outside Europe as well, i.e., Facebook and Google, that are targeting EU users.

Furthermore, the data protection regulation states a couple of new rules that will provide greater protection of personal data [27]. These rules are listed below and are to be incorporated within 2018.

- The right to be forgotten
- Better control over who holds one's private data
- The right to switch one's personal data to another service provider
- The right to be informed in clear and plain language
- The right to know if your data has been hacked
- Clear limits on the use of profiling
- Special protection for children

As we have seen, a lot of different aspects are combined to provide protection to individuals' personal information online. To be able to keep up with new technologies and services, both national and international authorities are continually working to better the regulations and laws concerning this topic. In Chapter 4, we will be looking at how a couple of the largest SNSs protect their collected user information and if they transfer user data in agreement with the presented laws.

# Chapter 3
# Methodology

This chapter describes the methods used to investigate the thesis' research questions and the reasons for the choices made. We will also take a look at the challenges and limitations regarding these methods.

The main goal of the thesis is to study to what degree users care about privacy on social media platforms. To be able to gain insight into this topic, we constructed the three objectives presented in Section 1.1. Consequently, we have utilised various research methods to answer each of them;

- Document Analysis
- Testing
- Quantitative Study

## 3.1 Document Analysis

Document Analysis is the method of reviewing and evaluating documents to receive a qualitative understanding of the analysed subjects [28]. The process of answering the first research question led us to perform a review and a comparison of documentation provided by SNSs regarding the protection of the users' information, e.g., privacy policies. By evaluating the different types of documentation given by Facebook, Google+, LinkedIn, and Twitter, respectively, we gained valuable insight into how large social media treat their users' information.

To structure the findings, we constructed the following sub-questions:

- What information is stored?
- How is information used?

- How is information collected?
- Where is information controlled?
- How long is information stored?

The questions were answered for each network in turn and included an additional section with remarks on clarity and language used in the policies. The results from this study are given in Chapter 4.

## 3.2 Testing

Conducting a practical testing allowed us to observe the effects of the practices described in the background study. The aim of the second research question was to gain insight into which companies, in addition to the chosen SNSs, that collect information about users. By performing tests on various websites and mapping the presence of third-parties, we got a better understanding of the extent of information sharing online.

Several free online tracking tools were considered for the purpose of mapping third-party trackers. Based on recommendations from the Norwegian Data Protection Authority, we decided upon Privacy Badger and Ghostery Browser Extension, which both display third-parties operating on websites. Additionally, Mozilla Firefox was chosen as the test browser.

To capture the extent of information sharing, we decided to run tests on 18 various websites in turn, in addition to the chosen SNSs. We found inspiration in Alexa's list [29] of top websites in Norway, and the selected sites were further categorised as either Norwegian News Sites, Norwegian Sites, or International Sites.

An additional goal was to find out if there are any correlations between results when accessing the sites being signed in as a social media user and when not. The tests were executed using the tools as follows;

1. Ghostery – not signed into any social media networks
2. Ghostery – signed into all chosen social media networks
3. Privacy Badger – not signed into any social media networks
4. Privacy Badger – signed into all chosen social media networks

Note that all cookies and other stored information were deleted between each test. The results from the practical approach are presented in Chapter 5.

## 3.3 Quantitative Study

A quantitative study is usually performed where the focus is on classifying findings and constructing statistical figures to explain what is observed. A common form is opinion-based user surveys concerned with understanding the behaviour of a large group of people [30].

The last research question and the previous studies led us to conduct a user survey mapping Norwegian social media users' knowledge on the discussed topics of this thesis. We distributed the survey using Facebook. The next section describes the survey's design whose findings are given in Chapter 6.

### 3.3.1 Design of Survey

Good practices for designing a survey include ensuring it is as short as possible, has a logical structure, and do not include ambiguous questions [30]. Our goal has been to follow these guidelines and create a survey interesting for social media users to answer.

As a part of the report "Personvern - Tilstand og Trender" from 2016, the Norwegian Data Protection Authority included a user survey focusing on what people think on the subject of surveillance economy and the online advertising business. Because of these findings being concluded a relatively short time ago, we decided to focus more on users' knowledge of the considered topics. This includes questioning what users know or think they know, about information sharing online and whether they are aware of how much information they provide to SNSs.

The survey consists of 16 questions separated into the following categories;

- General Information
- Use of Social Media
- Tracking Mechanisms and Sharing of Information

The user survey was conducted, purposefully, to gain insight into the respondent's knowledge concerning privacy in social media, and the questions is listed in Appendix B.

Additionally, the respondents were informed about the survey in general, the purpose, and that participation was both voluntary and anonymous. To conduct the survey, NTNU provided us with access to a service called SelectSurvey. As this service allows collection of personal data, we needed to acquire permission by the

Norwegian Centre for Research Data (NSD) to legally distribute the survey. The information sheet is given in its entirety in Appendix A.

### 3.3.2 Participants

The participants of the survey were, as mentioned, recruited using Facebook and our aim was to get a general representation of Norwegian users. Consequently, the targeted participants included all social media users in Norway, ranging between 13 and 80+ years old. The lower limit of 13 years old is set because this is the age limit for the majority of SNSs.

## 3.4 Challenges and Limitations

Challenges related to writing a thesis in twenty weeks may include time management and the restriction of scope. Additionally, the topics considered in this thesis are highly relevant nowadays, and we found constructing unique objectives to research difficult.

The process of meeting the thesis' objectives requires us to combine several research methods and, accordingly, challenges may arise with each of them. Parts of the documentation study relies on the opinions and comprehension of the authors, and, therefore, the results may be somewhat biassed. Difficulties may arise when performing the practical testing as well. We do not know whether the selected tools are reliable in detecting and reporting on third-party trackers.

The process of designing the right type of questions for a survey may prove to be a challenge. Especially when the goal is to map people's knowledge, i.e., extract the right kind of information, and at the same time ask unambiguous questions.

Additionally, there are challenges related to the distribution of the survey. Using social media could result in the loss of control of the participating group. Consequently, this could end up with respondents ranging in the same group, all having the same educational level, or not receiving enough responses. The latter case may happen if the survey is only available for a short amount of time.

When combining several methods, some challenges may arise when concluding the thesis. Discussing and presenting findings from various research questions in a suitable manner may both be time-consuming and challenging. Therefore, it is important to see enough time being set aside for this.

# Chapter 4
# Comparison of Privacy Policies

Facebook is the largest Social Networking Site (SNS) in Norway. The network has more than 3 200 000 Norwegian profiles, and 3 192 000 of them uses the service frequently [4]. Google+, LinkedIn, and Twitter follow Facebook with just above 1 million Norwegian profiles each [4]. On the whole, we have just over 6 million different SNS accounts in Norway.

As discussed in Section 2.4, a privacy policy is a document explaining how and why websites collect, use and manage user information. Even though a website provides a privacy policy, it does not mean that they protect personal, or any, information. Reportedly, this is a common misinterpretation among Americans online, where more than 50% believe a privacy policy ensures information to be kept confidential. This is according to a survey conducted by Pew Research Center [31] which further states that the average user rarely reads the privacy policies provided. Results for Norwegian users is given in Chapter 6.

Though, the numbers suggest that more than 6 million privacy policies should have been read in Norway, equalling 23 497 million words (calculated with numbers from Table 4.1). By reading at a rate of four hundred words a minute, the Norwegian population as a whole would spend 58 742 500 minutes reading privacy policies. However, these numbers only include privacy policies and most SNSs include additional policies and terms to give complete information about their services.

On May 24th, 2016, the Norwegian Consumer Council executed a live reading of the privacy policies of 33 apps found on an average Norwegian telephone. The live show went on for almost 32 hours, as can be seen from the screenshot in Figure 4.1. The purpose of the reading was to demonstrate the "scope, length and complexity" of the terms and conditions for digital services, and that reading these are an impossible task for most people [32].

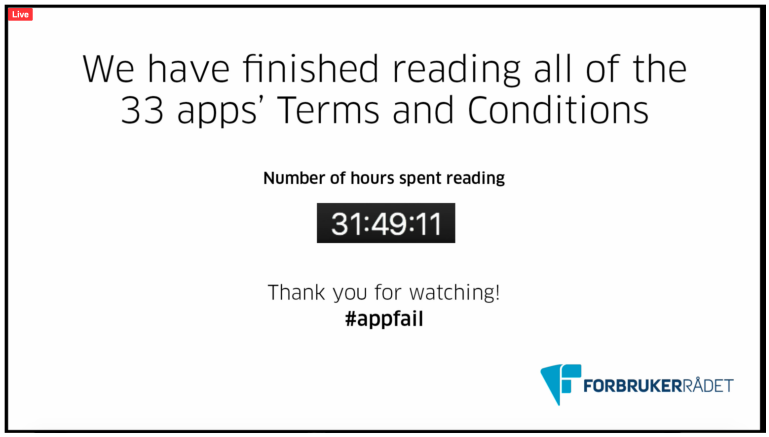


**Figure 4.1:** Screenshot from Completed Reading

All in all, studies argue that people do not read privacy policies, and the council claim that it be an impossible task. Still, websites provide privacy policies to give information about how they use the information they collect. SNSs often define their terms the actions performed on their Services. Consequently, we have made a table displaying some of the most used terms of this thesis' considered networks, how they compare to each other, in addition to a privacy policy word count. The information in given in Table 4.1.

### Evaluation Criteria

In this chapter, we are studying and comparing information from the privacy policies that applies to Facebook, Google+, LinkedIn, and Twitter. An evaluation of the services is conducted based on answers from the following questions;

- What type of information is stored?
- How is personal user information used and shared?
- How is information collected?
- Where is information controlled?
- How long is information stored?

As we remember from Section 2.1.1, data from SNSs can be divided into different types; Service Data, Disclosed Data, Entrusted Data, Incidental Data, Behavioural

**Table 4.1:** Social Networking Sites - Terms

| | **Facebook** | **Google+** | **LinkedIn** | **Twitter** |
|---|---|---|---|---|
| People that users form contact with | Friends | Circles | Connections | Followers |
| Acknowledge to have seen or like something someone else posts/shares | Like/React | +1 | Like | Like (heart symbol) |
| Where all new updates from the network are displayed | News Feed | Stream | Activity Stream | Timeline |
| Share something someone else has posted | Share | Share | Share | Retweet |
| Approximate number of words in Privacy Policy | 2 716 | 3 934 | 8 150 | 3 265 |

Data, and Connection Data. We use this categorisation (with subgroups) when considering collected data as stated in the respective privacy policies.

We will now look at each SNS in turn and consider each on their own before comparing our results.

## 4.1 Facebook

Founded in 2004, Facebook started as a networking site for Harvard students. It quickly escalated to include other universities, and Facebook has had a rapid growth from it was available to the public in 2006 [33]. In Norway, however, the service did not begin to expand until 2007. In, just a couple of months, the Norwegian user base increased from 3 000 to 80 000 [34].

Facebook is considered the largest SNS with more than 1 billion users worldwide, and over 3 million in Norway today. Ipsos states that 66% of the Norwegian population use Facebook on a daily basis [4]. From the categorisation in 2.1, we remember that Facebook is an example of a Social Network, meaning a network where users connect to others with similar interest. Hence, Facebook's mission is to "give people the power to share and make the world more open and connected" [35].

The main source of revenue is generated from advertising [36], and this allows

Facebook to provide a free-to-use service to their users. The users pay with personal data, as was mentioned in Section 2.3, including how they interact with the Service, and Facebook and third-parties provide the users with targeted advertising. We will get back to this later in this section.

Facebook offers a variety of products and services, including communication and advertising platforms. Their data policy applies to all of these additional, collectively referred to as Services. In the following sections, we are considering different aspects of Facebook's data policy. At the time this thesis was written, the date of the data policy's last revision was January 30th, 2015 [37].

### What information is stored?

Mandatory Service Data required for setting up a Facebook account is a first name, last name, mobile phone number or e-mail address, date of birth and gender. All other data is voluntary, though Facebook encourages people to submit additional information to improve their profile. The Voluntary Service Data includes a profile picture, hometown, school/university, religious/political view, among other things.

Facebook collects information from people users add as friends, users' relationships, and pages and groups the users "Like". Other information users provide in the contents they publish is in the form of text, photo, video, i.e., Disclosed Data. All this information can either be posted as public, visible for friends only, to a specified group of friends or restricted only to the user himself/herself, in a group or on a page. In addition, Private Communication Data is collected and associated with the respective accounts.

Incidental Data includes all the posts other Facebook users have published on the Services about the respective user. This information is not directly considered to be a part of a user's account, but Facebook collects and links it with the information already collected about each particular user.

Facebook also collects information on how users interact with the Services, i.e., Behavioural, Location, and Device Data. Depending on permissions users give, Device Data may include device identifiers, specific locations, mobile phone number, and IP address.

### How is information used and shared?

According to their data policy, Facebook collects different types of user information to be able to develop customised experiences and provide and support a consistent service [37]. Facebook states that people use their services to connect and share with

others [37] and that they make this possible by sharing user information with others in different ways.

First of all, information is shared with other people using the Services. The extent of this is dependent on the visibility settings of the shared content. The settings can be, as mentioned, set to public or restricted after the user's wishes.

Secondly, third-party partners and customers, e.g., advertising partners, receive information from Facebook. However, the data policy clearly states that only non-PII is shared. Advertisers receive information about users' age, gender, location, and device(s), giving them the ability to generate tailored advertisement [38]. Under "Privacy Settings", users have the option of denying that such information is to be used for this purpose. However, Facebook partners up with data brokers, whose functionality is described in Section 2.3, meaning that users might still be subject to targeted ads. We will come back to this later in this section.

Third-party websites often make use of Facebook's Social Plugins or lets visitors sign in to their service using Facebook credentials. Such third-parties may receive information such as age range, country/language, username, user identification, friend list, in other words, the user's Public Profile, as well as any information respective user agrees to share. In addition to third-party partners and customers, information is shared with Facebook's family companies [37]. These companies are, however, subjects of their privacy policies, and it is not specified in Facebook's data policy what kind of information is shared with them.

Finally, cookies and similar technologies, e.g., web beacons and fingerprints, may be placed on Facebook's site by third-parties. These third-parties include service providers, advertising partners, and so on. As mentioned, Facebook generates most of its revenue from advertising and these partners are therefore vital.

#### How is information collected?

Most of the information Facebook collects is derived directly from users and their behaviour when using the Services (e.g., signing up for an account, communicating with others, make a purchase in a game), i.e., Behavioural Data.

Another important source is other Facebook users, i.e., Friends, and the information they share to the Services. Including, sharing photos of other people, communication in groups, events, and messages. If any Facebook user imports his or her address book from their device, Facebook collects and stores all this information as well. Meaning that Facebook may have a user's' telephone number associated with their account even though the respective user has not provided this information themselves. Consequently, users have little control over how much information about

them Facebook collects as it is difficult to control what friends share.

Included among Facebook's sources of information are third-party sites that use Facebook's widgets, or so-called Social Plugins. These plugins are integrated with simple HTML code, as we can see in Figure 4.2 [39], and following is a list of some of the different options [40]:

- **Like Button** - Lets users automatically share content from third-party sites on their own Facebook profile, so that their friends can see them.
- **Share Button** - Let users share content on Facebook with particular friends, in a group or private message.
- **Send (on Messages) Button** - To let users share content from third-party sites privately to their friends.
- **Embedded post** - A "window" where public posts are visible into the content of a third-party site.
- **Embedded Video Player** - To display Facebook videos on third-party websites.
- **Page Plugin** - To embed components of a Facebook Page on third-party sites.
- **Comments Plugin** - To let users leave comments on third-party websites using their Facebook accounts.
- **Follow Button** - Lets users subscribe to others public updates on Facebook.

```
<script
        src="https://www.facebook.com/assets.php/en_US/sdk.js" async></script>

<div class="fb-like" data-href="{your-link-url}"></div>
<div class="fb-post" data-href="{your-post-url}"></div>
```

**Figure 4.2:** HTML Code for Facebook's Like Button and Embedded Post

A social plugin can collect information such as users' IDs, the websites they are visiting, and other browser-related information. It is used to improve Facebook's products and show people "more interesting and useful ads" [39].

As mentioned, Facebook partners up with data brokers [41] to help with tailoring advertisement. Based on preferences a Facebook user might have, the data broker match advertisers with people they would want to reach, as explained in Section 2.3.

The data brokers have cookies on different websites on the web and collect information about user behaviour. The collected information is later sold to Facebook. As mentioned, users may turn off the functionality that allows Facebook to use their information for advertising purposes. However, this setting does not apply to the Data Providers who might still tailor ads based on the user's preferences. Information on how users can completely opt out is not in the scope of this thesis.

**Where is information controlled?**

Facebook was founded in the US and has its main office in Menlo Park, California. They have expanded to have 14 offices in North America and 49 offices in total around the world. An overview of Facebook's offices around the world can be seen in Figure 4.3. All these offices could potentially access the information a user provides to the Services. The information collected is stored in Facebook's data centres which are located in the US and Sweden. Information is commonly stored in the data centre closest to the respective user.

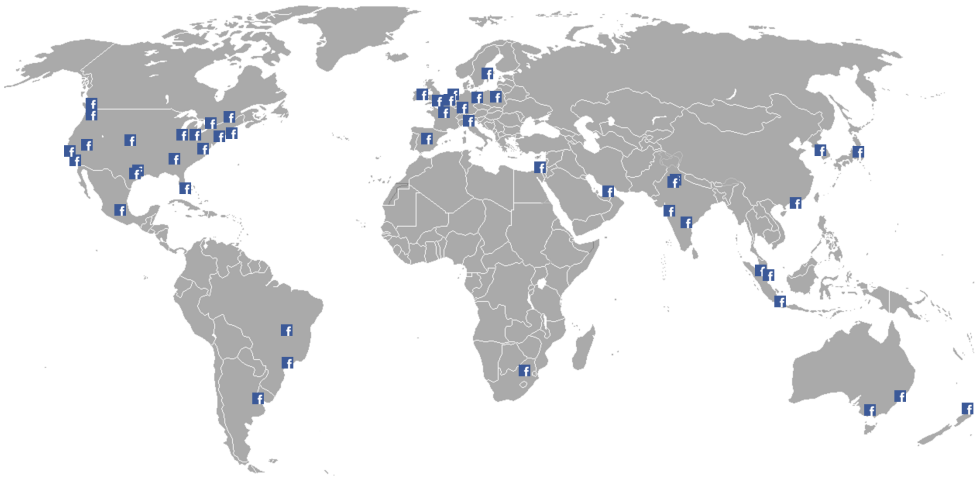

**Figure 4.3:** Facebook Offices

In Facebook's Statement of Rights and Responsibilities we find that "[users] consent to [have their] personal data transferred to and processed in the United States" [42]. Recalling from Section 2.4.1, there are strict regulations considering the transferring of personal data out of Norway, and out of the EU. To help the process, amongst other things, Facebook uses a data controller in Ireland for all users outside the US and Canada. A data controller is responsible for ensuring that the process of personal data complies with the PDA regulations [43].

In addition, "Facebook, Inc. complies with the US-EU and US-Swiss Safe Harbor framework" [37]. From Section 2.4.1, we remember that the Safe Harbor framework has been considered invalid since October 6th, 2015. Facebook's data policy does not state that they have taken any other action since this date, but from Safe Harbor's website, we find that Facebook's certificate runs out on October 5th, 2016 [44].

### How long is information stored?

According to Facebook, they store user information for "as long as is necessary" [37] and for as long as user accounts are active. However, if a user's information is needed to "provide products and services" [37], it still will not be deleted. Meaning that Facebook collects everything a user does and uploads to the Services and keeps this on their servers for as long as the respective account is active. This includes every wall post and photograph, every message sent, and a lot of tracked information about users' interactions on Facebook.

When or if a user chooses to delete his/her Facebook account, it may take up to 90 days to delete everything the user has posted [45]. It is important to be aware of the fact that information other Facebook users have shared about the respective user, in addition to content the user have shared to, e.g., groups, are not considered to be a part of the account in question. Meaning that the information will be available even after the account is removed. Also, messages sent between two people are not deleted until both participants in the conversation have deleted the messages.

### Comments

Facebook's data policy is written in a way that it is easy to understand. The layout makes it easy to read with different coloured categories and icons. When it comes to content, however, we find it to be quite vague. The word "may" is frequently used and the information given is superficial. For example, it is not clearly explained what kind of information is collected from or shared with Facebook's family companies. To get this information, users need to read the respective companies' privacy policy.

Additionally, the policy often states to collect information "such as" and gives a couple of examples. This leaves the users with an idea of the information gathered but does not clearly provide information about everything the service collects.

On the positive side, the data policy alone is relatively short with just over 2 700 words and the language and wording makes it easy to read. One could argue, however, that this compromises transparency. How much information is collected from e.g. social plugins? We do not know for certain, but it is assumingly large amounts.

Also, Facebook's Terms of Service, i.e., Statement of Rights and Responsibilities, refers to several others terms and policies in addition to the data policy, which makes the full documentation quite comprehensive. Ideally, to be aware of all the information provided, all these documents should be read in light of each other.

## 4.2 Google+

Google+ was launched late 2011, as a replacement for Google Buzz, and the goal is to connect people from diverse backgrounds and have them form new communities and share interests [46]. In 2015 the Service went through a redesign as an effort to narrow the scope of the network.

On official lists, Google+ is listed with more than 2.5 billion users, but still, Facebook is considered to be the largest SNS [47]. The reason for this is the relatively low rate of monthly active users which is approximately 300 million. In Norway, for example, the service has 1.2 million users but only 10% of users over 18 years old use the service on a weekly basis [4]. An explanation for these numbers is that for every Gmail address, an account is automatically set up on Google+ as well.

Google+ is free-to-use and is, as mentioned, set up for every Google Account. Google can provide this service for free because they collect information about their users. As long as users are signed into any of Google's services, they are passively using Google+, and thereby providing Google with collectable information [48]. Meaning that users are paying for the service with personal information.

Google has a joint privacy policy for all Google services. Some of them, like Gmail and YouTube, are subject to their own privacy policies in addition to the one, but that is not the case with Google+. The information in the following sections is thereby based on the joint privacy policy, in addition to other documents concerning different aspects of Google+. At the time this thesis was written, the privacy policy was last modified on March 25th, 2016 [49].

### What information is stored?

If a user wants to create a Google+ profile, they cannot do so without setting up a Google Account. This, on the other hand, gives access to all Google Services. The Mandatory Service Data required is first name, last name, username, password, date of birth, gender and location. Voluntary Service Data includes a mobile phone number, current email address, and similar.

Google+ allows people to post texts, photos, links, videos, events and polls which constitute Disclosed Data. These posts will then be associated with that particular user and the user's account. The posts can be published publicly, to a circle or

a specific user. The Entrusted Data may include recommendations, i.e., +1's, or comments on other users posts, and re-sharing of posts. Posts other users provide about a user or re-sharing of his/her posts is not associated with that particular user's account.

In addition, Google Services collect much additional information such as Behavioural Data and Connection Data. As mentioned, users are passively giving information to Google when they are signed into the services. However, Google states that once information is associated with the users' accounts, Google treats it as personal information [49]. Google Services is a big business and, consequently, it is not only the data from users' behaviour on Google+ that is associated with their account.

**How is information used and shared?**

Google primarily use information to "provide, maintain, protect, and improve" [49] their many Services. The name provided in a user's profile is used across all services Google offer that require a Google Account. Google says that they may collect and combine information about users from their various services, including personal information [49].

Google's privacy policy states that Google will display a user's profile name, profile photo, and Behaviour Data collected on Google or third-party applications connected to that particular account [49]. Including the use of names and pictures in commercial contexts. Information utilised in the context described above is what Google calls "shared endorsements" [50], and may be used in advertisements. What this entails is that if a user reviews or recommends something from a third-party website using their Google+ profiles, then both name and profile picture is visible by default. The consequence is that it may turn up in different advertisements on other services. An example of this is shown in Figure 4.4. Users do, however, have the option of changing their visibility settings to prevent this if they want to.


**Figure 4.4:** Shared Endorsements by Google

One of Google's well-known products is Google Analytics. Google Analytics is used, by both Google and other companies, to improve user experience by using collected user information from cookies, pixels, and similar [49]. They do this by, for example, displaying tailored advertisement. The privacy policy clearly states, however, that no sensitive information, e.g., sexual orientation, race, religion, and health, is associated with any identifiers from cookies or similar technologies.

As mentioned Google may combine and share personal information across their services. For example, if a user performs a search on Google's search engine, while being signed into a Google account, the search would result in not only results from the public web but also photos, pages, Google+ posts, and similar, from friends and people they follow on Google+ [51].

Google´s privacy policy is not clear on how, what type, or how much information they share with third-party websites. However, it does say a lot about how it is collected, which we will take a closer look at in the following section.

### How is information collected?

Google collects massive amounts of information from all of their Services. Consequently, Google's "file" on each user probably contains an unimaginable amount of information. The information people share on Google+ is stored in addition to all the already collected information from use of Google search, Gmail, among others.

Google's privacy policy states that the reason for the collection of information is to provide better services to all of Google's users [49]. Meaning, figuring out things such as what language people speak, the kind of advertising people respond to, or which connections online means the most to people.

The number one source of information is the users themselves, the information they provide, and how they use the Service. This includes the +1's of, for example, articles or photos to other users, content shared in communities, conversations, and how they interact with Google's services outside the Google platform. More on this in the upcoming sections.

As mentioned, Google uses information from cookies, pixels, and similar technologies, to enhance their Services. These methods are also used to collect information from any site that uses Google's advertising services, any of their social plugins, or from their partners [49]. In a video, explaining how Google uses cookies, Maile Ohye, Senior Support Engineer at Google, states that "most of the time, there is no personally identifiable information in a cookie file" [52]. She does not, however, say anything about when PII is enclosed in a cookie.

Third-party websites can implement any of Google+'s plugins. These are integrated in a similar way as Facebook's social plugins, as we remember from Section 4.1. Google receives information about anyone who uses the buttons or visits the sites they are on. Google's social plugins include [53];

- **+1 Button** - Similar to Facebook's "Like Button". This allows users to recommend content from third-party sites to their circles.
- **Google+ Badge** - Let users find Google+ profiles, pages, or communities, in a "window" on a third-party website.
- **Follow Button** - Let users add people or sites to their circles without leaving the third-party site.
- **Google+ Share Button** - To let users share content of a third-party site to their circles.

Third-party websites that use these buttons are not allowed to try to discover the identity of any user who uses them. The exception is when a user uses their Google credentials to sign into a third-party application, and consequently allows the third-party access to their information. Further, the third-party websites are not authorised to sell or transmit any user information related to the user's use of any of the buttons. Including the use of pixels, cookies, or other similar technologies [54].

**Where is information controlled?**

The Google Services control the information associated with Google+. Google has 18 offices in the United States and 60 international offices. This is illustrated in Figure 4.5. Google's data centres are located several places in the United States, but

also in Taiwan, Singapore, Ireland, the Netherlands, Finland, and Belgium. Meaning that personal data associated with Google+ accounts can be found in any of the Google office locations.

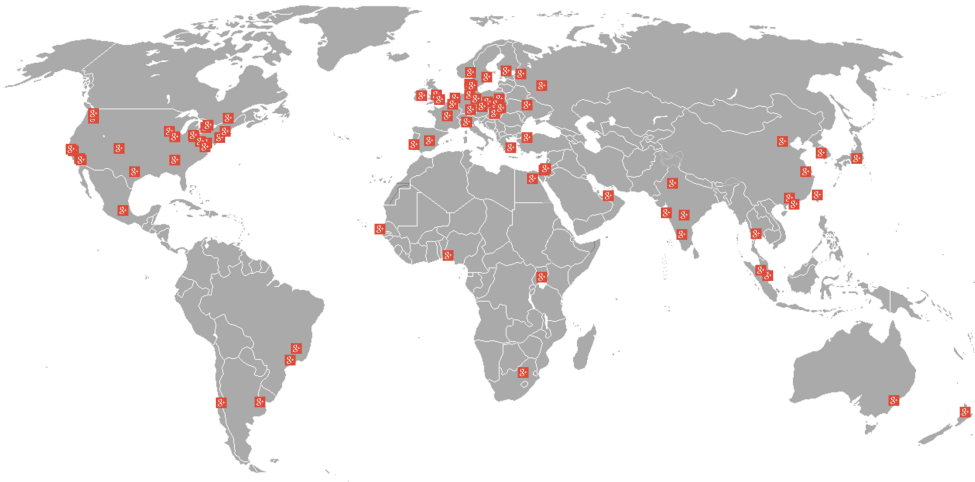

**Figure 4.5:** Google Offices

As Google has data centres located in the EU, data from European users is primarily stored in the closest data centre, and thereby does not need to be transferred outside the EU. Google Ireland Limited controls that the data transfer between the EU and the US happens in agreement with the PDA [55].

Google complies with the US-EU Safe Harbor Framework and the US-Swiss Safe Harbor Framework regarding the collection, use and preservation of personal information from EU member countries (and Switzerland) [56]. However, this information, stated in Google's Self Regulatory Frameworks, has not been updated since 2014, and there is no information about the Framework not being valid anymore.

### How long is information stored?

When a user signs up for a Google Account, Google states that they keep the basic information the user provides. In addition, when the user is signed into his/her Google Account, Google stores and protects anything the user creates using Google's Services so that users "will always have [their] information when [they] need it" [57]. From this, we assume that Google Services stores the information users provide for as long as they can.

To delete a Google+ profile, the user has to sign into a separate Google website and perform the required actions [58]. The information associated with the user's Google+ profile will thus be deleted. Information that is related to other users' accounts or public information on Google+ will be disabled or "hidden from other users", or even remain public on the Services. In addition, photos uploaded to Google+ will, for example, not be deleted as this is considered to be part of a different service, i.e. Google Photos [59].

We have not been able to find precise information about the duration of the deletion process at Google+. For Google Apps, we have found that Google will delete data from their system "as soon as reasonably practicable within a maximum period of 180 days" [60]. Still, we have to assume that even though a user deletes their Google+ account, Google will still control a lot of their personal data. This because their Google account will remain active and all information about the user is associated with this account.

**Comments**

Google+ do not have an own privacy policy or Terms of Use. Google's privacy policy applies to all Google Services, and given the amount of information these collect on a daily basis, we find their policy both partial and vague. Also, the policy links to a large number of additional sites and documentations that provide users with significant amounts of information about the different Google products, apps, and so on. It is not clear what applies to which products and, thereby, it is not easy to find precise information regarding Google+.

Google's privacy policy is divided into sections with headings, but as the entire documentation is in black and white, it is perceived as a lot to read. The text also includes sentences that are underlined and by hovering over these, the user is provided with an example and a link to additional information about the topic. As the users have to read other documentation to understand what is meant by terms stated in the privacy policy, we find Google's privacy policy quite extensive.

The privacy policy consists of 3 934 words and by itself does not provide much accurate information. "We may" is frequently used and there is a lot of what we find pretty vague information. This comes from the wording used in the policy and includes "information like" and "information such as". Google does provide some examples in these cases, though the precise information is not given.

## 4.3 LinkedIn

LinkedIn is a business-orientated SNS where the goal is to connect professionals from all over the world [61]. The company was founded in 2002 and has since grown to

become an important tool in the business world with 414 million active users [62] on a global basis. With over 1.2 million registered profiles, LinkedIn is recognised as one of the largest SNSs in Norway [4].

Users set up profiles that include education and employment history. They can add professional connections and follow companies, and information from these connections is shown on a professional network news feed. A user's network is made up of "1st-degree, 2nd-degree, 3rd-degree connections" [63]. More on this in the upcoming sections.

LinkedIn has three primary sources of revenue, namely, talent solutions, marketing solutions, and premium subscription products [64]. Talent solutions make up around 55% of the total revenue while marketing solutions, for example, advertising, create about 25%. The remaining revenue comes from users choosing one of four premium accounts. These "provide members with better access to contacts in the LinkedIn database" [65].

LinkedIn's privacy policy also applies to two other services, i.e. Pulse.me and SlideShare. In the next sections, we will be identifying different aspects of the privacy policy and other documents concerning a user's personal information on LinkedIn. The policy was last revised on October 23rd, 2014, at the time this thesis was written.

#### What information is stored?

To become a member of LinkedIn, users have to provide their real name (first and last name) and email address, which constitutes the Mandatory Service Data, in addition to a secure password. Only one account is allowed per user, and this is why the user's real name must be provided [66].

Voluntary Service Data may include a mobile phone number, postal code, job title, company name, skills, professional experience, educational background, and so on. LinkedIn encourages users to add as much information as possible as this will give the user better experience using the service. A user can post and comment on others users posts, and LinkedIn provides a messaging system between users as well. All of this information is collected and associated with the user's account.

LinkedIn collects information about users' connections, and how they interact with them and the Service. Additionally, LinkedIn receives Device Data and Location Data. Some of this information is associated with the users' respective personal accounts. Other information, for example, some interactions with other users and in groups are sometimes considered public information.

### How is information used and shared?

All collected information is used to help LinkedIn provide a viable and sustainable service for their users [67]. For example, contact information, i.e., Service Data, is used to communicate with users, e.g., send service messages, newsletters, and invites, and information such as Behavioural or Incidental Data can be utilised for service development, e.g., provide customised experiences and develop new features.

Information is, naturally, shared with other people using LinkedIn's Services, including non-members. As mentioned in the introduction, a LinkedIn user's connections are listed at different levels. For example, a 1st-degree connection can view the user's full profile and contact information while for a 2nd- or 3rd-degree connection information might be restricted. All of this is dependent on the user's privacy settings.

LinkedIn's privacy policy states that they share personal information with third-parties with the user's consent, and where it is necessary to carry out the user's instructions. Further, it is used when essential to providing features and functionality to the user, when the law or other legal processes require it, or when it is necessary to enforce the user agreement [67]. Personal information may be shared with LinkedIn's affiliates, e.g., LinkedIn Corporation may share with LinkedIn Ireland, when it is reasonably necessary to provide the Services [67].

Information collected from users' interaction with LinkedIn's plugins and cookies on third-party websites is used to tailor advertising and develop personalised functionalities, among other things. LinkedIn may create reports to the third-party sites hosting the technologies, based on the collected information [67]. The privacy policy clearly states, however, that no personal data is enclosed in these reports.

### How is information collected?

Information is first and foremost derived from the users and how they use the Service. This includes both mandatory and voluntary Service Data. In addition, users have the option of syncing contact lists, calendars, and similar services with LinkedIn. This information, including all phone numbers, is stored and used to help users "manage and leverage [their] contacts in connection with [LinkedIn's] Services" [67].

Behavioural Data from both users and their connections is collected and associated with user accounts. Users provide collectable information every time they click on an advertisement, perform searches, or when they comment on or share a post.

LinkedIn offer plugins for third-party websites with simple customised JavaScript code. The plugins LinkedIn provides are listed below [68].

- **Share Button** - Let users share content to their connections at LinkedIn.
- **Follow Company** - Let users follow the updates and posts of a company.
- **Member Profile** - Displays a member's profile on a third-party site, with or without connections, and the option of establishing contact with that member.
- **Company Profile** - Displays a company's profile on a third-party site and the possibility of following that company.
- **Company Insider** - Shows how many of a user's connections are employees at a given company.
- **Jobs You Might Be Interested In** - Displays available job positions at a given company.
- **Alumni Tool** - Displays where former students at a given school are currently working, live, what they do, and similar. How this looks for NTNU is shown in Figure 4.6.

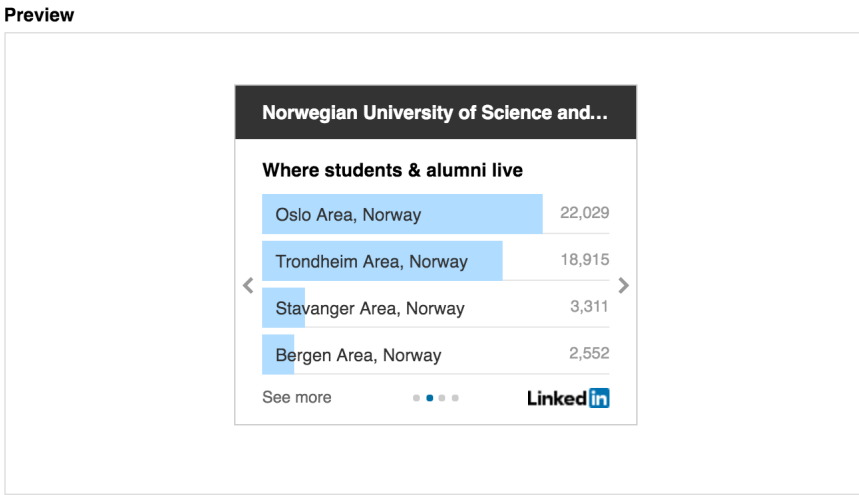


**Figure 4.6:** LinkedIn's Alumni Tool

Any website that incorporates one or more of these plugins sends information to LinkedIn and a user's behaviour is associated with that particular user's account [67]. Additionally, users sometimes have the options of signing in on third-party services using LinkedIn credentials. By doing this, they allow LinkedIn to receive information about their actions and content they view on these sites. Additionally, they allow the particular website insight into the information they have provided to LinkedIn.

LinkedIn receives information from third-party cookies and uses it for advertising and analytic reasons. This information is also used to help with talent and marketing solutions [67]. Some of the third-party cookies that LinkedIn allows are Google Analytics, DoubleClick, Eloqua, and BlueKai [69]. On the positive side, LinkedIn provides its users with links directing to sites where they can opt-out of third-party cookie services [69]. Also, the privacy policy states that LinkedIn will not place cookies through plugins in browsers that belongs to non-LinkedIn members.

### Where is information controlled?

LinkedIn has eight offices in the US and 22 international offices. These can be seen in Figure 4.7. We have not been able to find information directly from LinkedIn about their data centre infrastructure. Other sources, however, claim that they operate two data centres in the US and are in the process of expanding with two more, one more in the US and one in Singapore [70].

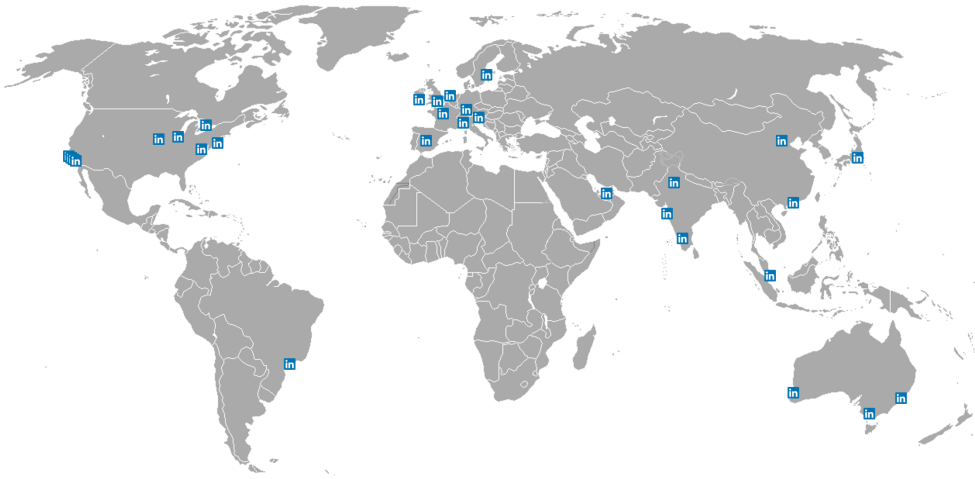

**Figure 4.7:** LinkedIn Offices

It is stated in the privacy policy that LinkedIn may transfer and process user information outside the user's country of residence, to wherever LinkedIn operates [67]. By this, we assume that a user's information could potentially be sent to any of the 23 countries the Service is located. All user information is stored in either the US or Singapore. Remembering from Section 2.4.1, LinkedIn need to provide EU/EEA users personal data with extra protection when transferring their data.

LinkedIn state that they comply with the US-EU and US-Swiss Safe Harbor Frameworks regarding the collection, usage, and retention of personal information from EU member countries and Switzerland, in addition to adhering to the Safe Harbor Privacy Principles [67]. A notice has also been added to the policy informing users of the fact that the Safe Harbor Framework is no longer recognised as a legal information transferring mechanism from the EU to the US.

In addition to the note in the privacy policy, LinkedIn informs their users about the issues at hand and which mechanisms LinkedIn now use to protect the transferred data. Also stated is that "the text of the Privacy Shield remains a work in progress and is subject to formal approval, so LinkedIn monitors for the agreement details and the data protection authorities' interpretation and reactions" [71].

#### How long is information stored?

According to the policy, LinkedIn holds on to users' personal information as long as their accounts are active or as necessary to provide their services [67]. LinkedIn keeps all the information a user publishes to their account to provide the best possible service. They also keep track of all posts, comments, and messages, and we assume this is for as long as it is not deleted.

LinkedIn removes data from the Services within 24 hours of the closing of the account. However, users should be aware the deletion process and de-personalization of any logs or backup information can take up to 30 days. Information shared with other LinkedIn users, or information others have copied, is not considered a part of a user's personal account and will remain visible on the Services.

#### Comments

LinkedIn provides their users with an 8 150-word long privacy policy, making it quite comprehensive. However, most of the information a user needs is provided in this documentation and only has a few sites to complete their documentation. The layout is presented clearly with lines separating the different aspects. The privacy also includes a summary related to each section.

The privacy policy is written in a way that is easy to understand, and the users are provided with explanations concerning different aspects of the policy. However, the policy also includes vague statements as “we may” and “we attempt to”, and in many cases do not provide the users with precise information.

On the positive, LinkedIn clearly informs its users about opt-out possibilities regarding cookie use. Even though this will not restrict the amount of information LinkedIn receives, it limits what is shared with third parties and other connections on the Services.

## 4.4 Twitter

Twitter is an online SNS, often referred to as a microblog, as mentioned in Section 2.1, that lets users publish and read short 140-character messages called “tweets”. The Services had 305 million active users in the last quarter of 2015 [72] and just over 1 million of these were Norwegian.

Founded in 2006, Twitter originated as a place for people to create and share ideas instantly and without barriers [73]. Some use Twitter as a news feed by following famous people, businesses or networks. Others use the service as a sort of “private” chat room by limiting their followers and the people they follow to close friends and family. Lastly, people also use it as a microblog for updating their followers about their daily lives.

Twitter is a free-to-use service and generates about 85% of its revenue from advertising. The rest is derived from Data licensing, meaning that Twitter sells tweets daily to companies for analysis [74].

Twitter’s privacy policy applies to any user registered to the Twitter Services, i.e., to publish Tweets, and users of any of Twitter’s other services, e.g., TweetDeck, Curator, Digits, and Periscope [75]. Collectively, these are referred to as Service from now on. In the next sections, we consider information surrounding Twitter’s collecting and use of user data. The privacy policy was last updated on January 27th, 2016, at the time this thesis was written.

**What information is stored?**

The Mandatory Service Data required when creating an account on Twitter is a name, either a real name or a pseudonym [74], e-mail address and password. All other Service Data provided by the user is voluntary and can include username, a phone number, biography, location, date of birth, and similar. The user can also import their address book to the Service.

The Service's main feature is for users to be able to share information in the form of Tweets. These can include text, photo, video, and links, and are by default posted publicly. Retweets are also considered a part of users' personal information, meaning repost someone else has posted. Additionally, Twitter has a private messaging feature and the content of these messages is associated with the respective users' accounts.

Incidental Data includes tweets followers, or others, tweet about the user and private messages. If a user deletes a Twitter account sent messages will be removed, but messages other users have sent will remain in the Service as they are considered a part of the other user's account.

Twitter also collects a lot of other information about each user. Including interactions with links across the Services and information from cookies on third-party websites. Additionally, Twitter collects Connection Data [74].

**How is information used and shared?**

Information and content are first of all shared with other people visiting or using Twitter. By default, Tweets and Twitter profiles are public, meaning that people without a Twitter account can freely view and search user profiles. However, Twitter users have the option of updating their privacy settings, such as making tweets protected, i.e., only visible for users' followers, or add/remove location information [74].

Twitter does not require their users to provide a real name. Meaning that users have some control of how much Twitter knows about them, concerning PII. There are actions, however, that users should think carefully about before performing if they want to remain "anonymous". For example, connecting other social media accounts, e.g., Instagram or Facebook, to their Twitter accounts. Resulting in information sharing between companies and thereby more data about users is accessed by Twitter.

Similarly, users may want to use their Twitter credentials to sign in to third-party services or websites. Users should be aware that Twitter, in this case, shares information, such as phone number, with that party [74].

Information collected from cookies, local storage, and similar technologies are used to deliver and measure the Service in different ways. For example, help users

log into Twitter, personalise the content they see, save their preferences, or show them relevant ads [74].

### How is information collected?

Twitter receives much information directly from user input and activities, i.e. Service Data and Behavioural Data. Specifically mentioned is the fact that users have the option of, for example, importing their address book from their phone to the Service. This information is thereby stored and may be used to tailor content.

Information is collected by third-party partners and affiliates as well, through widgets and cookies. Twitter uses these technologies on their sites and services, and on other websites that have integrated any of Twitter's plugins [76]. Twitter's social plugins options include:

- **Tweet Button** - Lets users share content from a third-party site directly as a tweet.
- **Follow Button** - Allows users to follow the third-party website's or service's updates and tweets on Twitter.
- **Hashtag Button** - Lets users tweet stories with hashtag (Tweet #TwitterStories).
- **Mention Button** - Allows users to tag a given person or company, and similar, in a Tweet.

Twitter uses other third-party services for information collection to improve their Services and measure and tailor advertisement. These services, for example, Google Analytics, collect information about users and share this on Twitter. This information may include, browser cookie IDs, websites visited (in the form of URLs), or information about users' devices.

### Where is information controlled?

Twitter is an American company and, consequently, administrates their Services from their headquarters in San Francisco. The company has 13 different offices in the US and 28 international offices spread around the world [73].

Figure 4.8 shows where Twitter offices are located in the world. The information a user provides to the Services could potentially be located in any of the markings on the map. It is also defined in the privacy policy that Twitter may transfer or store any user information to any of the countries Twitter operates [74].

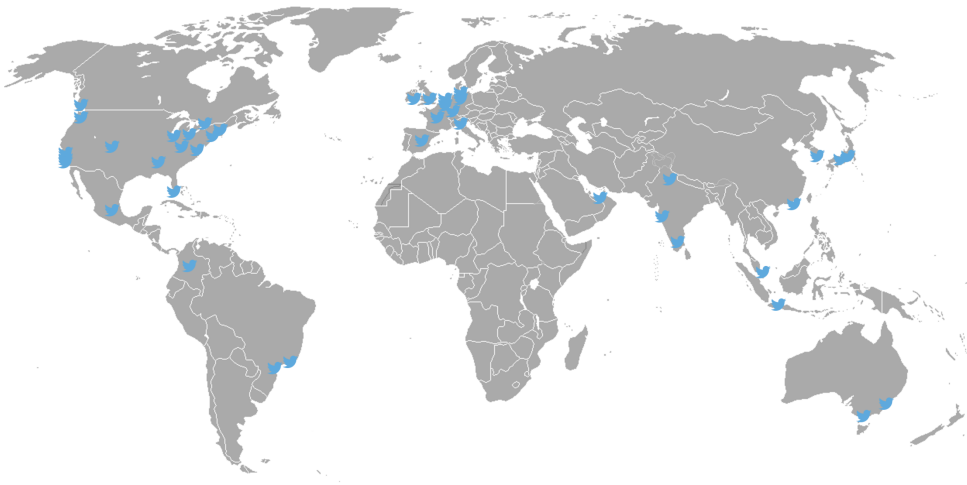

**Figure 4.8:** Twitter Offices

A data controller controls the information for Twitter users outside the US. The company responsible is, in this case, the Irish company Twitter International Company. Twitter's privacy policy also includes that the users control and are responsible for any information they post on the Services.

Twitter does not provide any information about which mechanisms are used when transferring data from the EU/EEA to the US. From Section 2.4.1, we recall that there are strict regulations to protect the EU citizens data. On Safe Harbors website, we find that Twitter has a current Safe Harbor certification and that Twitter agrees to "[c]ooperate and [c]omply with the EU and/or Swiss Data Protection Authorities" [77].

### How long is information stored?

Twitter does not delete any of the information a user publicly provides to the Services, e.g., Tweets. The privacy policy clearly states that the company's default is almost always to make the information users provide public for as long as they do not delete it [74]. If users want to remove the information, they have to delete each tweet physically. The process of completely remove the information can take some time, but will eventually be deleted from the Services.

If users wish to delete their accounts, they must first deactivate their Twitter account as the Service does not provide an actual "Delete account"-function. Thirty days after the date of deactivation, Twitter will "begin the process of deleting [users']

account[s] from [Twitter's] services" [74], and the process takes up to a week.

As mentioned before, information about users may be present on the Services that is not associated directly with their accounts. This information remains on Twitter after accounts have been deleted, and will stay on the Service for as long as the user who published the content is an active Twitter user.

**Comments**

Twitter's privacy policy holds just over 3 200 words, the language is vague and the word "may" is used a lot. For example, Twitter is unclear when it comes to what type of information is shared with or received by their corporate affiliates, stating that "[Twitter] may also receive information about [users]" [74].

Additionally, Twitter uses the phrases "like" and "such as" frequently in their policy, concerning information collected. They do, however, often provide many examples when these are used. Still, the use of "information such as" does not give precise information about what type of information is collected.

On the positive side, Twitter honours the Do Not Track browser option and also warns its users to be careful with what they post using Twitter's Services. As we remember from earlier in this section, minimal Mandatory Data is required to be able to use the Service. Therefore, much responsibility lies with the users.

When it comes to transparency, Twitter provides a site dedicated to this topic. Users can request different reports on issues where, for example, Twitter have been legally ordered to disclose account information or remove content [78]. However, this site is not referred to anywhere in the privacy policy or in the Terms of Service, which we find odd.

## 4.5 Comparison

We have in this chapter given insight into four different privacy policies to figure out how SNSs protect the user's personal information. We will now summarise our findings.

All in all, we see four differences between our SNSs. Namely,

1. Mandatory Service Data
2. Data the users provide/share (Disclosed and Intrusted Data)
3. Privacy settings
4. Deletion time

The differences stated do not directly concern the protection of personal information, but they are in accordance with how much information the SNSs are protecting.

There are significant differences in how much information is required for setting up accounts in our chosen SNSs. The reason for this is, we believe, due to the nature of the SNSs, but also the differences in how users intend to use the Services. Facebook, Google+ and LinkedIn are all originally made to connect real people and acquaintances. On Twitter, however, the objective is for people to share their ideas and interests, either anonymously or not.

**Table 4.2:** Mandatory Information Comparison

| | **First Name** | **Last Name** | **Gender** | **Email** | **Phone number** | **Date of Birth** | **Location** |
|---|---|---|---|---|---|---|---|
| **Facebook** | ✓ | ✓ | ✓ | ✓(or phone number) | ✓(or email) | ✓ | |
| **Google** | ✓* | ✓* | ✓ | | | ✓ | ✓ |
| **LinkedIn** | ✓ | ✓ | | ✓(or phone number) | ✓(or email) | | |
| **Twitter** | ✓** | ✓** | | ✓ | | | |

* "Google+ profiles are meant for individual people". Google, therefore, recommends users using both their first and last names on their profiles [79].

** Twitter does not require real name; a pseudonym is enough.

Table 4.2 displays the Mandatory Service Data needed to set up accounts in the different SNSs. As we can see, Facebook and Google+ require the most information and recall that the information applies for a Google Account. Both Facebook and LinkedIn require users to provide their real names to the service and both state that users are only to have one account. Google, on the other hand, only encourage users to add full names and to limit themselves to one account. Twitter, on the contrary, has no restrictions concerning this and allows both name and pseudonym.

The SNSs all encourage users to provide the service with additional information to the Mandatory Service Data. Here, the differences in how much users can include vary. On Facebook users can add just about anything, even sensitive information such as political view and religion. All the chosen networking sites allow users to publish posts that, potentially, can include any information. This is out of the networks' control as, in the end, the users themselves decide what to share to their circles or groups of friends or connections.

We also find differences in how users themselves can protect the information they share through the services, i.e., differences in privacy settings. Though, the nature of the various SNSs affect how information is shared between users. On both LinkedIn and Twitter, the purpose is to reach a large number of users, and thus, information is often public.

How information is collected and controlled is the same for all four networks. They all use the same technologies for collection, e.g., widgets and cookies, and use the same policies for protection, e.g., Safe Harbor principles. However, LinkedIn is currently the only network which informs its users about the Safe Harbor framework being invalid.

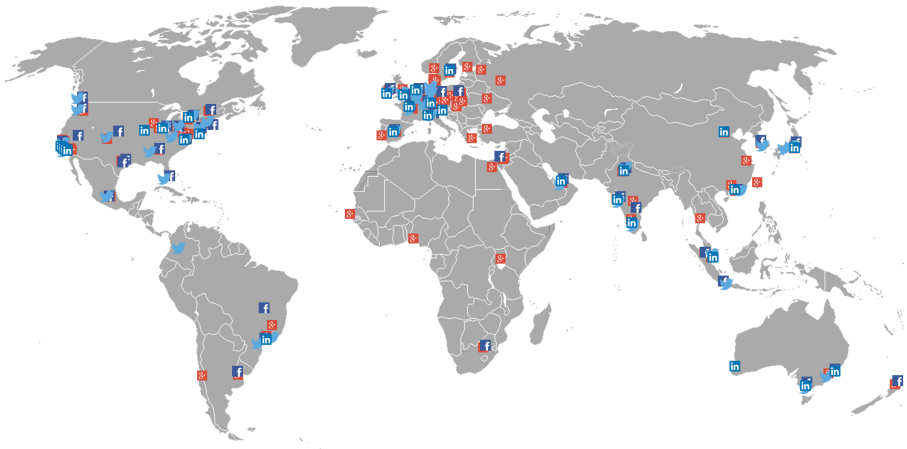

**Figure 4.9:** Social Networking Site Offices Around the World

Each of the discussed SNSs are American companies. They control their Services from main offices in the US in addition to having a European office located in Ireland. Also, by agreeing to each SNS's privacy policy and Terms of Service, users permit the networks to transfer and process information in any country the respective company operates. As we can see in Figure 4.9, there are many similarities of the SNSs office locations.

Another similarity is that all four of the SNSs store information as long as they possibly can. The reason for this is to provide the user with the best experience by ensuring that users can access their shared and posted information at any time. The differences arise when considering deletion of accounts. The duration varies from 24 hours to 90 days depending on the SNS in question. A common factor is that they delete all the information considered to be a part of the user's account. The information users have shared publicly or to groups, and other members will remain on all the different Services.

# Chapter 5
# Mapping of Third-Party Trackers

The aim in this chapter is to gain insight into the websites tracked by the SNSs analysed in the thesis. We have included a mapping of third-party companies tracking users across various websites, how many trackers there are, and which companies appear on multiple websites. By combining this information, we gain insight into which companies collect information about users when browsing the Internet and, additionally, provide an illustration of the complexity of information sharing.

To understand what a third-party tracker is, we must first examine what happens when we visit an arbitrary website. Recall from Section 2.3.1, the explanation of what happens regarding information exchange and real-time bidding during advertising trading. In addition to this, different components are loaded on the website to provide the users with the right content. These components may be provided by other domains and servers than initially requested, i.e., from so-called third-party sites.

A website owner's goal is often to provide content relevant to a specific group of audience. To do this, the owner may need resources to provide the different services. For example; if the owner wants to support video on the website, there is a need for a video player, or if the owner wants the content to be free for users but still wants to make money, an advertising partner may be of interest. The result of all this is that a lot of different companies can potentially watch what visitors are doing, and consequently, map their behaviour. Below is a list of how some third-party services can be integrated to websites, and Figure 5.1 provides a visual example of this.

- Tailored advertising, e.g. using cookies.
- Site analytics, e.g. using cookies.
- Widgets from social networks, e.g. social buttons, or forecast companies.
- Video players, e.g., Flash Video Player and YouTube plugin.
- Embedded images, e.g., from Flickr.


**Figure 5.1:** Third-Parties on vg.no

To be able to give an answer to the second research question, we are performing tests on some of the most popular websites in Norway. In the following sections, we provide a short introduction to the analytics used and the websites we have chosen to test. Following, we explain how we executed the testing and present the results.

## 5.1 Analytics

To be able to map the third-parties tracking behaviour online, we chose to use two free available software products and compare the results. Both tools utilised are recommended by The Norwegian Data Protection Authority as indications to who is tracking users online. The following sections present the chosen tools.

### 5.1.1 Ghostery Browser Extension

Ghostery Browser Extension shows users all the companies that are tracking them when while visiting different websites. It is developed by Ghostery, Inc., a global marketing technology company whose goal is to provide online transparency and control for individuals and businesses [80]. To do this, Ghostery, Inc. offers a variety of tracking and marketing tools. We will in this thesis be using their free browser extension, hereby referred to as Ghostery.

#### How does it work?

Ghostery's main functionality is to monitor all web servers that are being called when visiting a website. These servers are then matched with Ghostery, Inc.'s library of data collection tools, i.e., trackers [80]. If a match is detected, that tracker shows up in a pop-up information bubble or a control panel in the browser.

Ghostery looks at the websites HTML code to see if there are any "tags" or "trackers" placed on the site by third-party companies. The library of trackers contains enough information to be able to tell if the tracker is placed for advertising purposes, if it is collecting data, or adding functionality on the site [80].

What differs Ghostery from ad-block plugins or manually deleting cookies is that Ghostery blocks all communication between the browser and the third-party web server. Opting-out or blocking cookies will stop tracking, but allow altered communication between the parties.

### 5.1.2 Privacy Badger

Privacy Badger is a browser extension by the Electronic Frontier Foundation (EFF). This is a non-profit organisation whose goal is to defend "civil rights in the digital world" [81]. Privacy Badger's main purpose is to block advertising and track cookies that do not respect the "Do Not Track"-setting as mentioned in Section 2.2.4.

#### How does it work?

Privacy Badger tracks all third-party domains that embed scripts, images, and advertising on websites. As mentioned, a third-party server may track users without users' permission. When this happens, Privacy Badger will automatically block content from that third-party. Privacy Badger can detect technologies such as uniquely identifying cookies, local storage super cookies and canvas fingerprinting [82]. Additionally, it recognises cases where third-party domains are responsible for important features, e.g., video players or embedded images, on the site. These connections are allowed, but tracking cookies and HTML referrers are blocked.

Privacy Badger contains a feature which causes social media widgets to be replaced with a "stand-in version" [82]. An example of this can be seen in Figure 5.2. This allows users not be traced by the SNSs unless they explicitly choose to click on the widget. It is important to note, however, that Privacy Badger will not replace social media widgets unless the associated tracker is blocked. If real widgets appear on the websites, this means that Privacy Badger has not detected tracking variants from the widget or that the website has implemented their own version of the widget.


**Figure 5.2:** Stand-In Version of Facebook Widget

## 5.2 Test Sites

In addition to the four chosen SNSs, we wanted to research which companies are tracking users on some of the most popular websites in Norway. This was done, necessarily, to gain insight into the extensive network of information sharing happening on the Internet. Alexa Internet, Inc. provides lists of websites, both globally and by country, based on traffic [29]. From Alexa's list of top sites in Norway, we chose 22 different websites and ordered them into four categories. These can be seen in Table 5.1.

**Table 5.1:** Categorisation of Websites

| **Social Networking Sites** | **Norwegian News Sites** |
|---|---|
| facebook.com | vg.no |
| twitter.com | nrk.no |
| linkedin.com | dagbladet.no |
| plus.google.com | tv2.no |
| | aftenposten.no |
| | nettavisen.no |
| **Norwegian Sites** | **International Sites** |
| finn.no | youtube.com |
| dnb.no | msn.com |
| yr.no | wikipedia.org |
| startsiden.no | netflix.com |
| difi.no | reddit.com |
| sparebank1.no | ebay.com |

## 5.3 Execution

In all, we completed four tests by using the tools listed in Table 5.2. Each session was performed using a newly installed Firefox browser, resulting in no cookies, bookmarks, or any information being stored before commencement. We continued to access all test sites and each categorisation in turn. Starting with Social Networking Sites, continuing to Norwegian News Sites and Norwegian Sites, before ending with the International Sites. We followed the order of sites shown in Table 5.1 from top to bottom.

We deliberately chose to run the tests with Ghostery and Privacy Badger separately as the tools affect each other. As we remember from the previous section, Privacy Badger replaces real social media widgets with stand-in versions. Meaning that Ghostery will not pick up these trackers as Privacy Badger has already blocked them.

**Table 5.2:** Testing Tools

| **Type** | **Name** | **Version** |
|---|---|---|
| Browser | Firefox | 46.0.1 |
| Analytic | Ghostery | 6.2.0 |
| Analytic | Privacy Badger | 1.7.0 |

Firstly, we visited the sites with Ghostery installed and without being signed into any of the SNS. Cookies, or other information, were not deleted between the sites. All third-parties were registered for each site before moving on to the next test. We then signed into all four SNSs, with a private account, and followed the procedure once more. When finished, Firefox was reinstalled, and we completed the same procedure with Privacy Badger. We followed the order listed below.

1. Ghostery – not signed into any SNS
2. Ghostery – signed into all SNS
3. Privacy Badger – not signed into any SNS
4. Privacy Badger – signed into all SNS

Next section will present the results of the tests. Including, the types of companies that appear on the different sites and their uses. Additionally, we will give insight into a couple of the dominating companies found and their role in this context.

## 5.4 Results

In the following sections, we have merged the results from the different tests we conducted. Firstly, we will present our findings with regards to the chosen SNSs, i.e., on which websites these companies appear. Then, we continue to analyse all findings from the other categorisations from Table 5.1 in turn.

#### Social Networking Sites

A summary of our results displaying which of the 22 test sites the different SNSs appear on is shown in Table 5.3. In this listing, Google consists of all Google and DoubleClick products including, but not limited to Google Analytics, Google Adsense, and DoubleClick Floodlight. Similarly, the Facebook column includes all Facebook and LiveRail products.

Table 5.3 shows that Google appears on and collects information on over 80% of the test sites. We also see that Facebook is present on 9 of the 22 websites. However, the information from Table 5.3 refers to the social networking companies, and not the actual SNSs.

We did, however, find Facebook Connect present on four different sites, namely dagbladet.no, startsiden.no, sparebank1.no, and msn.com. Facebook Connect allows users to use their Facebook credentials across the Internet. Meaning that this service will send information about the respective Facebook user to the company that has implemented Facebook Connect. Additionally, we found Twitter Button, as described in Section 4.4, on msn.com.

Finally, the results showed Facebook Exchange on vg.no, aftenposten.no, msn.com, and tv2.no. Facebook Exchange (FBX) is the ad exchange concerning advertising shown on facebook.com. This means that Facebook tracks their users across the Internet to provide them with targeted ads once they enter their services.

Other than the findings presented above, our results were not clear on when the other SNSs were present on the websites and did not show any results concerning these SNSs directly.

With regards to what parties are present on our SNSs, we find that both Facebook and Google only allow their products to track users. On the other hand, LinkedIn and Twitter allow Google products, DoubleClick and Google Analytics, respectively, as well as their products. All in all, SNSs allow very few third-parties on their sites.

**Table 5.3:** Social Networking Companies on Other Websites

| | **Facebook** | **Google** | **LinkedIn** | **Twitter** |
|---|---|---|---|---|
| **facebook.com** | ✓ | | | |
| **plus.google.com** | | ✓ | | |
| **linkedin.com** | | ✓ | ✓ | |
| **twitter.com** | | ✓ | | ✓ |
| **vg.no** | ✓ | ✓ | | |
| **nrk.no** | | ✓ | | |
| **dagbladet.no** | ✓ | ✓ | | |
| **tv2.no** | ✓ | ✓ | | |
| **aftenposten.no** | ✓ | ✓ | | |
| **nettavisen.no** | | ✓ | | |
| **finn.no** | | | | |
| **dnb.no** | | ✓ | | |
| **yr.no** | | ✓ | | |
| **startsiden.no** | ✓ | ✓ | | |
| **difi.no** | | ✓ | | ✓ |
| **sparebank1.no** | ✓ | ✓ | | |
| **youtube.com** | | ✓ | | |
| **msn.com** | ✓ | ✓ | | ✓ |
| **wikipedia.org** | | | | |
| **netflix.com** | ✓ | ✓ | | |
| **reddit.com** | | ✓ | | |
| **ebay.com** | | | | |
| **Total** | 9/22 | 18/22 | 1/22 | 3/22 |

### Norwegian News Sites

Some news websites, on the other hand, allow a much higher number of third-party trackers on their websites. Our results show an average of 60 trackers on both vg.no and dagbladet.no. The majority are related to advertising, while a few are recognised as site analytic tools. Nettavisen.no allows a significant number of third-parties as well, approximately 40. Nrk.no is the news website we analysed that allows the least amount of trackers. Our results show only five third-parties present, all of them being site analytics because nrk.no does not provide advertising.

The third-parties present on Norwegian news websites are mostly a mix of American and European companies. Google Analytics, for example, is used to track all six websites. Several of them are also utilising Linkpulse, which is a Norwegian analytics company. When it comes to advertising, the companies vary from site to site. Doubleclick and Adform appear on all websites, again with the exception of nrk.no.

**VG example**

We wanted to see if there are any differences between what appears on the front page of a news site and what appears directly in news articles. Because of this, we conducted an additional experiment on vg.no. The results are presented in Table 5.4 and show clear differences when cookies are deleted, and when they are not, before accessing the article.

**Table 5.4:** Differences VG.no

| | **With deletion of cookies** | **Without deletion of cookies** |
|---|---|---|
| **VG-article** | **Advertising** (35): Adform, Admeta, Advertising.com, AppNexus, Audience Science, BidSwitch, Criteo, DataXu, DoubleClick, Facebook Exchange (FBX), Improve Digital, Index Exchange (Formerly Casale Media), LifeStreet Media, Lijit, LiveRail, Media Innovation Group, Media.net, MediaMath, myThings, OpenX, OwnerIQ, PubMatic, Quantcast, Right Media, Rocket Fuel, Rubicon, ShareThrough, SMART AdServer, SpotXchange, Taboola, Teads, TNS, TripleLift, Turn Inc., Videoplaza | **Advertising** (11): Adform, AppNexus, Audience Science, DoubleClick, InSkin Media, Integral Ad Science, Lotame, Moat, TNS, TubeMogul, VideoPlaza |
| | **Site Analytics** (2): AT Internet, New Relic | **Site Analytics** (2): AT Internet, New Relic |

Firstly, we accessed vg.no's front page and navigated to an article acting like a "normal" user, i.e., by not deleting cookies in between the sites. Resulting in 60 third-party trackers on the front page and 13 on the article. Next, we repeated the process, but now with deletion of cookies in between the sites. This time, we found 61 trackers on the front page and 37 on the article. This shows an increased number of third-party trackers when deleting cookies, and other stored information, before entering the article. We believe this is caused by the fact that cookies have already

been set when accessing the article directly from the front page. When deleting cookies in between, we found that many of the cookies we registered on the article also were present on the front page. In other words, when visiting sites by the same domain, we believe cookies only need to be set once.

When social media widgets are present on websites they are passively collecting user information, as mentioned in Chapter 2. However, during the VG-experiment none of these were registered, despite the fact the vg.no article provides social plugins to both Facebook and Twitter, as shown in Figure 5.1. Facebook Connect and Facebook Social Plugins did not appear until the comment field at the end of the article was clicked. The reason why the plugins failed to be recognised is unclear. It may have something to do with the analytic tools used, e.g., the type of connections and communication they detect, or the implementation of the social plugins.

#### Norwegian Sites

The number of third-party trackers on the other Norwegian websites is quite low. These websites vary in area of use, but common for five out of six is that they allow no more than eight third-party trackers. Startsiden.no is the only one allowing a higher number and allows more than doubled compared to the other sites. Google is, again, present on each site in the form of either Double Click, Google Analytics or both. Otherwise, there is no correlation between the results.

#### International Sites

From the international websites, we expected a presence of many third-parties, but our results proved differently. YouTube, as a Google product, only have Google products present, while Wikipedia has none at all. Two websites stood out, consequently msn.com and ebay.com, and both allowed approximately 15 trackers each. Google are prominent in both cases, with DoubleClick.

## 5.5 Dominating companies

Our results show that some third-party companies appear more regularly on websites than others. The most common reasons for third-parties to be present on a website are for either site analytics or advertising. The following section provides a closer look into some of these companies, given in Table 5.5 to help gain insight into the roles of third-party trackers on websites.

**Table 5.5:** Dominant Companies

| Category | Company |
|---|---|
| Site Analytics | Google Analytics |
| | TNS |
| | Linkpulse |
| | New Relic |
| Advertising | DoubleClick |
| | Adform |
| | AppNexus |
| | AudienceScience |

### Site Analytics

The goal of a site analytics company is to optimise and understand how people use the Web. They accomplish this by measuring traffic on websites, by collecting and analysing behavioural data [83]. The following sections include information about the four most prominent companies in this category from the test results.

**Google Analytics** tracks and reports website traffic, and is a common web analytics tool on the Internet. Tracking of sites commences by customers of the service implementing a block of JavaScript code on their websites. When the code is executed, information about a visitor's browser and computer settings is collected [84]. Additionally, the script is configured to set cookies in the visitor's browser which gather information about the current session, among other things [85]. Information collected by Google Analytics is owned by the customers, i.e. the companies which have implemented the service. Google Analytics' Terms of Service states that the service is not to be used to collect PII [86].

**TNS Gallup** is a well-known market analysis bureau in Norway. The company analyses markets in various ways, for example, by tracking traffic on Norwegian websites. TNS uses cookies to analyse visiting trends, and they emphasise that no PII is collected [87]. The information TNS Gallup collects is used within the company and not sold to any third-parties [87].

**Linkpulse** is another Norwegian company, and their analytic tool is the most popular choice for online news media in Scandinavia" [88]. They use a tracking script to set cookies and analyse Internet traffic in real-time. Linkpulse's analytics tool does not collect IP addresses, personal information, or browser type. This is true for their cookies, as well, because these do not store or collect personal information [89].

Also, Linkpulse state in their privacy policy that they do not share information with third-parties and that information gathered is the property of their customers [89].

**New Relic** is an American company that produces software analytics. Their technology monitors web and mobile applications in real-time, and allows customers to easily view and analyse massive amounts of data [90]. The trackers identified in our testing with relation to New Relic correspond to their browser monitoring product. These tracking cookies collect performance data only, and IP addresses are not stored [91]. New Relic processes the information they collect on behalf of their customers but do not have any relationship with the individuals behind the personal data [92].

### Advertising

From Section 2.3 we remember that many companies and different aspects are required to display advertising on various websites. The companies presented in the following sections provide websites with platforms to show advertising or provide the actual advertising.

**DoubleClick** is a subsidiary of Google, and the purpose of the company is to both provide and develop ad serving services on the Internet [93]. The technology helps their clients, for example, advertising agencies and media companies, to analyse advertising campaigns. DoubleClick cookies are used to improve advertising [94] and they do not store any PII. Also, information obtained from these cookies is never associated with information from other Google services [95].

**Adform** is a Danish company that provides customers with a digital advertising solution, and specialises in real-time bidding and programmatic media. By the use of this technology customers can collect non-PII about Internet users and thereby analyse how the users interact with their advertising [96]. Adform uses cookies and stores cookie-based profiles. The information stored includes, but is not limited to, operating system, geographic location, and URLs and facts about interactions with advertising [96]. Unlike most of the other companies discussed in the previous sections, Adform allows customers to collect PII but does not encourage them to use their technology for this purpose.

**AppNexus** provides customers with a platform to buy, sell, and deliver online advertising. This includes interest-based advertising and is often done through real-time bidding. The information these platforms collect and store, using cookies and similar technologies, is done over time and regards users' web browsers and devices across various websites and applications [97]. Also, it is clearly stated that no PII is collected from either the platform solution or AppNexus themselves.

**AudienceScience** delivers targeted or personalised ads to consumers by working with advertisers, publishers, and other businesses. They use cookies to create what they call "audience segments" which includes anonymous data collected across the Internet [98]. The cookies do not store any PII, but rather the information relevant for the advertiser regarding visitors' interests or any information that can be used to tailor advertising.

## 5.6 Summary

The results from the previous sections show that SNSs do not appear on many other sites than their own. The exceptions are Facebook Connect that appeared on four additional sites, and Twitter Button, on msn.com. On the other hand, we found Google products to be present on over 80% of our testing sites. The information collected might not be directly linked to Google+ user profiles, as is the case previously with Facebook and Twitter. However, Facebook also collects information by other means than Facebook Connect, i.e., LiveRail or social plugins. Resulting in much information being collected by the large SNS companies on the different websites.

Other results show that there are many other third-parties following users across the Internet. Tracking is usually done for advertising or site analytic reasons. In either way, third-parties do not collect or store PII. This, at least, applies to the majority of the companies we evaluated in Section 5.5. The information collected concerns mostly device and location, and by itself this information cannot identify the individual user.

The analysed companies from Table 5.3, state that they do not give their collected information to other businesses. The exceptions are to family companies and when the user explicitly gives permission. Businesses that own a SNS, such as Google, will, however, receive information from the respective SNS in addition to the company's other collected information. As Google products are present on both LinkedIn and Twitter, we assume they additionally receive some information about LinkedIn and Twitter users.

Figure 5.3 illustrates the complexity of third-party information sharing discussed in this chapter. Here, we can see the icons of the 22 tested websites and a random selection of third-party trackers discovered during our testing. The eight dominating companies, i.e., the companies with the most connections, are located in the middle of the figure. Note that Wikipedia has no connections in this illustration, and the reason for this is that we did not find any third-party trackers present on the website.

It is also important to note that the figure does not display the reason, e.g., advertising or analytic, for the third-party company to be present on the respective websites. It shows a simple illustration of how traffic of information flows across the web.

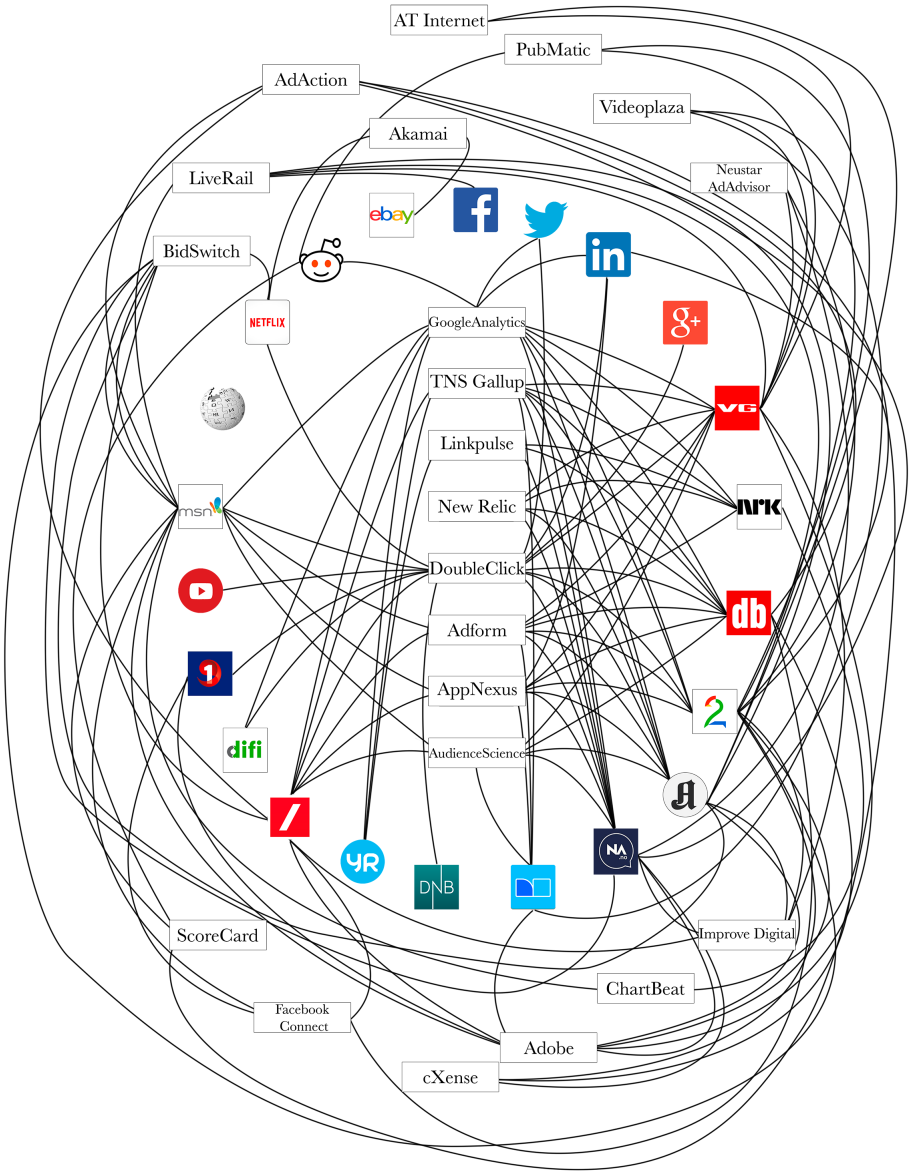


**Figure 5.3:** Mapping of Third-Party Trackers

# Chapter 6
# Test of User Knowledge

This chapter presents the findings from the user survey. Conducting a user survey has been an important goal for this master thesis as it is a valuable way of gaining insight into user knowledge and habits on Social Networking Sites.

The user survey was created on the basis of the findings from the previous documentation analysis and tests, in addition to the provided background study. By evaluating the information businesses provide to their users and understanding how information is shared online, we felt conducting a user survey was the best solution to determine the users' knowledge about the subjects.

We had the survey active for two weeks and received 526 responses. Unfortunately, some respondents skipped a couple of questions causing our findings to range from 523 to 526. We will, when necessary, provide comments on the total number of responses. An overview of the distribution of respondents given by gender, age, and education level is listed below.

- Gender (Total responses: 525)
  - Men: 212
  - Women: 313
- Age Groups (Total responses: 524)
  - 13-18: 30
  - 19-24: 223
  - 25-29: 196
  - 30-39: 17
  - 40-59: 50
  - 60-79: 6
  - 80+: 2

- Education (What are you studying now? Optionally, highest completed education) (Total responses: 525)
  - Primary School: 2
  - High School: 73
  - Bachelor's degree: 151
  - Master's degree or higher: 299

The upcoming sections contain our findings presented in three parts; Use of Social Media, Tracking Mechanisms and Sharing of Information, and a section including additional findings. A complete list of the questions from the survey is given in Appendix B.

## 6.1 Use of Social Media

As previously explained, we chose to distribute our survey using Facebook. Therefore, it came as no surprise that a 100% of the respondents are Facebook members. Additionally, more than 90% access the network several times a day. Further, we find that approximately 50% have accounts on each of the other SNSs though the activity rate on these is much lower. Of Google+ members, 75% state to never use the service while the majority, i.e., 43%, of LinkedIn members visit the site only a couple of times a month. An overview of the member numbers is shown in Table 6.1. As we can see, 119 respondents, corresponding to 22%, are members of all four SNSs.

**Table 6.1:** Members of Social Networking Sites

| **Social Networking Site** | **Number of Members** | **Percentage** |
|---|---|---|
| Facebook | 526/526 | 100% |
| Google+ | 245/526 | 47% |
| LinkedIn | 311/526 | 59% |
| Twitter | 268/526 | 51% |
| **All** | 119/526 | 22% |

Furthermore, we included a question asking how often respondents share information, in the form of text, photos, or similar, on each respective SNSs. Facebook remains the most prominent with the highest activity rate, and the majority of members, 57%, share information on special occasions. On the remaining three networks, less than 10% on average state to share information more often than "rarely". Of the respective Google+ members, 66% declare never to share any content.

On the topic concerning the visibility of the content shared on SNSs, 288 out of 526, consequently over 50%, claim to both know about and to have previously changed the visibility settings. When inquiring about the awareness of the amount of the personal information they provide the networks, we received an approximately 50/50 answer between; "I only provide what is necessary" and "I provide what I want to". The results from the latter question are shown in Figure 6.1.

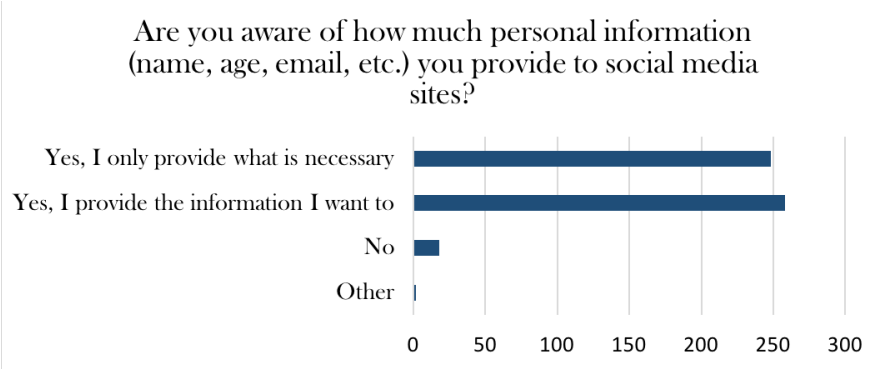


**Figure 6.1:** Amount of Personal Information

As we can see from the bottom bar in Figure 6.1, an "other"-option was included to the question. The reason for this was to allow the individual respondent to add a comment that was more representative to how they provide information to the networks. The added comments tell us that one person would rather not provide any information, and, consequently, only doing so when necessary. Additionally, another person reports to consider the situation carefully every time.

Figure 6.2 displays the distribution of responses concerning the visibility of personal information. 74 respondents, i.e., 14%, answers that they do not share this type of information online at all. These are all members of Facebook, 30 are Google+ members, 34 are members of LinkedIn, and 29 are members of Twitter. Hence, all 74 respondents have provided their name, i.e., personal information, to at least one SNS.

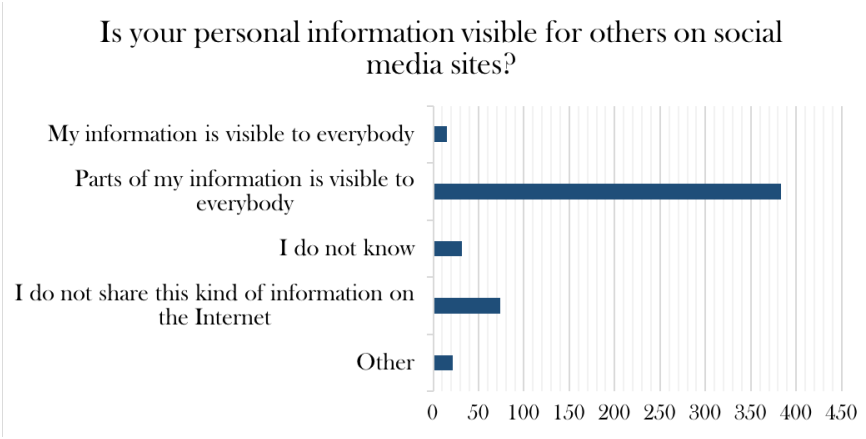


**Figure 6.2:** Visibility of Personal Information

## 6.2 Tracking Mechanisms and Sharing of Information

The following sections include findings from the last part of the survey, concerning the methods of tracking users online and the extent of information sharing.

We started off this part with a statement that reads as follows; "If a website has a privacy policy they protect your data". The findings show that 78% consider the statement as false while the remaining 22% believe it to be true. To investigate this further, we included a question about privacy policies, namely; "Have you ever read through a privacy policy?". Only 3%, i.e., 18/525, report to have read one thoroughly, 51% have never read one, and the remaining 45% claim to have skimmed through. The results from these two questions can be seen in Figures 6.3 and 6.4.

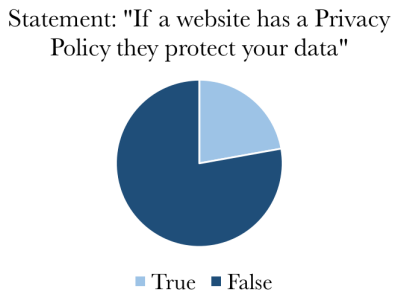


**Figure 6.3:** Privacy Policy Statement

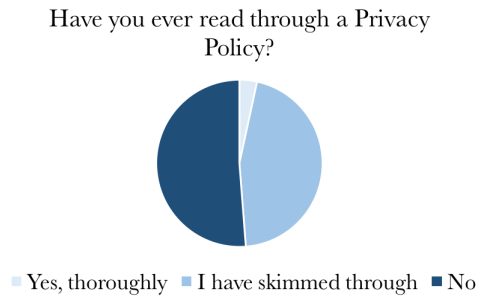


**Figure 6.4:** Reading of Privacy Policies

A closer investigation shows that all 18 respondents who have thoroughly read a privacy policy also answered the statement as "false". When only considering the respondents who believe the statement to be true, the majority, i.e., 65%, have never read a privacy policy, and no one has read one thoroughly.

Continuing on the topic of reading privacy policies, the respondents who have read one thoroughly all have an education of a Bachelor's degree or higher, and 78% have a Master's degree. In addition to answering correct to the statement of privacy policies not protecting their data, they proved to have a good understanding of why SNSs use tracking mechanisms. The majority knew that they are used to provide tailored advertising and to offer better services. However, 28% believe that trackers are used to giving third-parties as much personal information as possible. Additionally, 56% think that SNSs exchange information in the matter of receiving similar information from the third-parties.

Considering all respondents, the reasons for why SNSs use tracking mechanisms, and the purpose they serve, are well understood. More than 80% answer that SNSs use tracking mechanisms to earn money and show tailored advertising. Additionally, 57% reply that these mechanisms are used to offer better services.

When it comes to questions about cookies, the respondents show some knowledge on the topic. In all, 85% answer that "Cookies are used to map a user's activity (on a website)". Only 29% states that cookies are used to store username and password, while 19% believe they are used to encrypt information packets.

To map users' knowledge regarding the extent of information collection across the Internet, we included a question concerning Facebook's reach. Namely, "From which of these websites do you think Facebook collects information?". The provided alternatives included the tested websites in Chapter 5, i.e., the content from Table 5.1. We found no correlations between the answers received, but an average of 47% believe that Facebook collects information from the three other SNSs analysed in this thesis. However, 8% do not think Facebook gathers information from their site, facebook.com.

Lastly, we included a few questions on the topic of third-parties and their role on websites. On the question "How many third-parties, on average, do you think track what you do on vg.no?", we received no specific results, as shown in Figure 6.5. As we can see, the majority of the responses lie either in the range 6-49 or 90+. Only 10% of the respondents answered the correct range discovered in Chapter 5, which was 50-69 and an average of 60 third-parties present on vg.no.

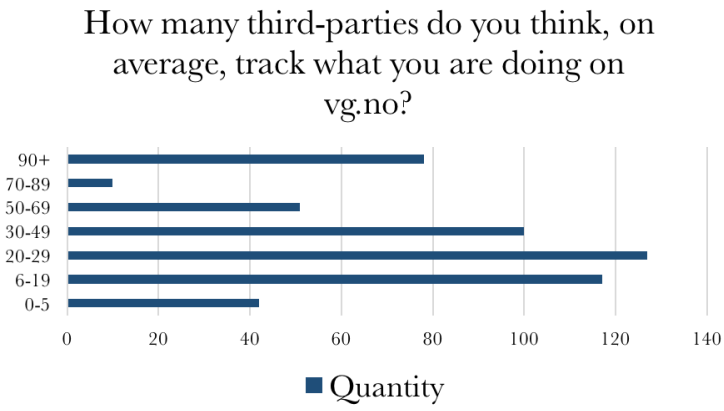


**Figure 6.5:** Trackers on vg.no

## 6.3 Additional Findings

The most prominent findings have been discussed so far. In this section, we take a closer look at different distributions of respondents, as mentioned in the introduction of this chapter, to see if there are any interesting variations.

We find that men and women answer in the same pattern on virtually every question, and, consequently, have similar percentage distributions. Still, there are some differences and these are found investigating the questions regarding the use of cookies and information collected by third-parties. 2% of the men state that they do not know what cookies are used for, while the numbers for women equals 13%. Additionally, 85% of the male respondents know that third-parties collect device information, while 68% of the women believe they do. As these are the most prominent differences, we can see that there are no major variations concerning the knowledge of tracking mechanisms either. Note that the percentages may be somewhat skewed because, as we remember from the distribution, there were approximately 60% female and 40% male respondents.

Furthermore, we compare findings according to the respondents' educational level. First of all, we find that the largest share of the LinkedIn members includes respondents with a Master's degree or higher. Approximately 34% with a Bachelor's degree or lower are members of LinkedIn.

Looking closer at a higher versus a lower degree of education, i.e., Master's degree and Bachelor's degree versus High School and Primary School, we find a couple of small variations. The responses to the statement, "If a website has a privacy policy they protect your data", show some uncertainty from the respondents with a lower

degree of education. This can be seen in Figure 6.6 which displays the comparison of the responses from the two groups. We also find a variation with regards to the reason for third-parties to use cookies. Only 7% of the respondents with a higher degree of education do not know why they are used, contrary to a 31% from the respondents from the other grouping.

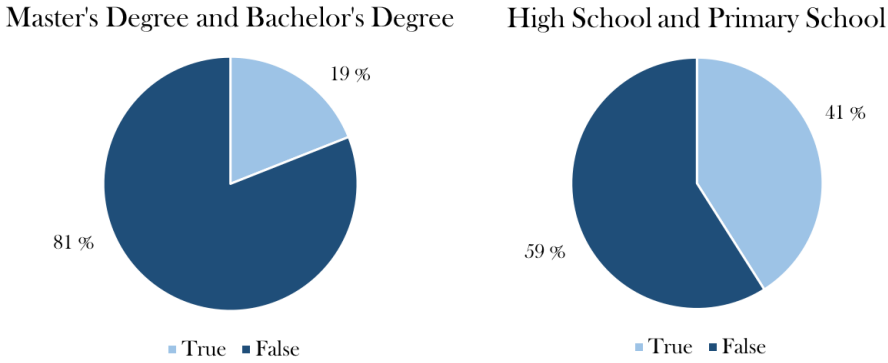


**Figure 6.6:** Comparison of Statement Responses

Finally, when comparing our youngest respondents, i.e., 13-29 years old, with the oldest, i.e., 30-80+, we find the main difference to be the willingness to provide personal information. The majority, corresponding to 67%, of the oldest group, report only to provide what is necessary. Contrary, the majority of the young group, i.e., 52%, state to provide what they want to. Other than this, we find no significant variations in the results when focusing on age groups. It is important to note, however, that this comparison consists of unequally distributed groups, as approximately 85% of the respondents are between 13-29 years old.

# Chapter 7
# Discussion

Throughout the thesis we have studied the content of several privacy policies, analysed third-parties on various websites, and conducted a survey to provide insight into user knowledge. We will in the following chapter discuss the findings from all previous chapters and relate this to the research questions. Lastly, we include a discussion of the limitations of the methods used.

## 7.1 Do social media networks protect their users in the same fashion or are there any differences?

The discoveries found with regards to the first research question turned out to be different from what we expected. We anticipated to find more detailed information regarding the protection of user data in the respective SNSs' policies. In Section 2.4, we briefly explained the purpose and content of a privacy policy. We quickly discovered, however, that the studied policies provide information in a vague manner, thereby requiring us to interpret the meaning of the statements provided.

As we recall from Chapter 4, required Mandatory Service Data and the options regarding privacy and visibility settings vary in each of the analysed cases. All other data users provide to the services are given voluntarily, and the different services enable users to share content publicly. This means that each SNS has different responsibilities regarding the protection of the collected information and to provide users with documentation regarding these topics accordingly. Our interpretation is that this is advantageous for the respective SNSs as they leave the main share of the responsibility of privacy to their users.

Technologies used for information collection by the various SNSs and the reasons for collecting data are essentially identical. As we recall from Chapter 4, the networks all claim to collect information to provide the best possible service to their users, i.e., using it to enhance and tailor the services. We have not found any precise information about what the collected data is used for or any clear statements disclosing how

these improvements are met. Leaving this matter up for interpretation, consequently, raises privacy concerns. We remember from Section 2.4 that all data collection must have a clear purpose, and we do not consider this to be upheld in the cases analysed.

Considering the language used in the respective policies also lead us to the types of data collected. We found the phrase "among other" repeatedly mentioned where the types of collected information are declared, and we find this worrying. What information does this really entail? The analysed SNSs state to use all the information they collect to improve the services and tailor experiences. Ultimately, this could entail using any information a user provides, i.e., including sensitive information such as chat messages and location data, to enhance services without the user's knowledge.

To summarise; all the considered SNSs provide their users with privacy policies and other documentation about the types of information they collect and how it is done. Included is also how the information is used and shared, where it is controlled, and for how long it is stored. However, we did not find the policies to give any precise information on how the networks actually protect the user information or the technologies in use for this purpose. This does not seem to be mandatory to include in policies, and we have, therefore, not been able to gain insight into this aspect.

We recommend users to practice caution regarding the information they share to SNSs and other services online, whether or not it is PII. Especially when using social media credentials to sign into third-party applications and websites. It is a convenient solution as it excludes the need to remember new usernames and passwords. However, the third-party company receives information from the respective SNS and collects additional information from the user's behaviour. Further, users need to adhere to additional privacy policies and documentation about how, and if, their information is protected by the third-party. A high level of understanding of what actions may lead to is, thereby, required among the users. If they do not comprehend this, they themselves are the greatest threat to their privacy online.

## 7.2 Other than the social media network itself, who else collects information about its users and how is the information spread between parties?

Our expectations when conducting the tests considering the second research question was to find the presence of numerous third-party companies. The results confirmed this and even showed more companies than expected tracking users, and thereby also SNSs users, online.

The thesis has described how companies gather information about users. The fact is that the combination of different types of information could give a clear picture of the respective individual, even considering non-PII. In Chapter 1, we touched upon the subject of Edward Snowden and that he got hold of and published sensitive information. When turning this around contemplating the companies analysed in both Chapter 4 and Chapter 5, one company stood out, namely, Google.

According to our findings, Google is the company with the furthest reach, and one might wonder how much information on the individual Internet user they possess. The tests show that their gathered information may include the activity from about 80% of websites visited, in addition to the information provided to Google+ and other Google products. Even though most Google products state to only collect non-PII, combining the information from their different services would most likely result in a clear picture of an individual.

Moreover, we expected to get some indication of what types of information are collected by the different companies online. Though, with limited time and the chosen tools, we were not able to obtain any results other than the purposes for the collection of information. When investigating SNSs and other companies tracking users online more closely, we found more vague information. Consequently, we still do not know precisely what information the different companies receive or share. A closer look at some of the dominating companies in Chapter 5 did not result in the wanted information either, as these also provide vague information.

Other than SNSs, our results show that a number of international companies, in addition to some Norwegian companies, collect information about Norwegian SNS users. This raises privacy issues concerning the topics presented in Section 2.4. Though the analysed SNSs in this thesis comply with the proposed regulations for transatlantic information transfer, we do not know to what extent other companies do so. We, therefore, do not have a full understanding of which companies are collecting user information and, additionally, not maintaining the adequate level of protection required by Norwegian privacy laws. This again means that the average user has little opportunity to gain knowledge on this matter.

As a result, we believe that no matter how much individuals are concerned about privacy, it is virtually impossible not to have one's information collected on the Internet. This means that users need to go to great lengths to stop companies tracking them. Fortunately, there are actions to enhance privacy online, some of which are mentioned in Section 2.2. Other options include making the use of incognito mode in browsers or connect to an anonymous network.

However, note that we have not found any tracking companies that gather or share PII. The important part is that users are aware of what information is collected, and the impact of their actions and what they are agreeing to.

## 7.3 What do social media users know in terms of how and how much information is being spread? Do they care?

Before conducting the user survey, our belief was that respondents with a higher degree of education would be more enlightened with regards to the subjects included in the questionnaire. This was not the case, and it proved that the level of knowledge was virtually equal between the groups. Where the respondents lack is in the understanding of the extent of information sharing online. The majority know they are being tracked, though only a few understand to what degree third-parties are present on websites. Findings also show that many have heard of and know about the technologies used online. What we understand from this is that there is still a long way to go before the general population fully understand the consequences and extent of the mechanisms used online.

We discovered that the average respondent has not read a privacy policy thoroughly, as was in line with our expectations. However, does this mean that only the 3% care about privacy and how their information is used online? We believe that the respondent chooses convenience over knowledge in these cases. The consequence is that most of them have accepted the SNSs' policies without thoroughly considering the potential privacy issues that may follow. How are they supposed to have an understanding of how information is collected and used if they do not read the provided documentation?

Fortunately, there is frequent media coverage addressing privacy related topics, for example when the Norwegian Data Protection Authority releases reports. The findings from the survey proved that people do have an understanding of how information used and the reason for its collection on SNSs. Meaning that the majority do receive information concerning the topics in some way or another, even though they do not deliberately read the provided documentation themselves.

We anticipated older respondent groups to be more sceptical than the younger respondents with regards to sharing personal information online, and we were proved right. The surprising fact is the frequency of which the younger respondents share content. This turned out to be rarer than we expected, and we believe the reason for this is the modern use of social media where people are focusing on staying connected rather than sharing. Different social medias are also specialising in customising individual services, and people are therefore using multiple services to meet their needs.

To summarise, we find the respondents to have a good understanding of tracking mechanisms online and why they are utilised. However, we find them to be moderately informed with regards to potential privacy issues of sharing information on SNSs. A small share of the respondents claims not to share personal information on the Internet while at the same time admitting to having an account on at least one SNS. Meaning that such information has already been shared. Accordingly, they do not understand the extent of sharing information online.

## 7.4 Privacy and Social Media: Do Users Really Care?

The authors of this master thesis have experienced that reading and understanding the contents of privacy policies is not only time-consuming but also challenging, as has also been proven by the Norwegian Consumer Council. Additionally, findings from the user survey show that the respondents have the same experiences, as only 3% read them. Though, Internet users claim to do so by checking off the "Yes, I have read and agree to the terms"-statement every day. As a consequence, the users do not have the adequate level of understanding of what they are agreeing to.

This fact is further substantiated by other questions from the survey. When questioned about if they usually change the visibility settings when they share content on SNSs, the vast majority either check the settings every time or have previously done so. This gives the impression that the respondents consider the reach of the information they provide the services. However, their behaviour says otherwise. Many users are willing to provide the information they want to the services, often including PII, which gives the impression that privacy might not be that important to the users after all.

Moreover, the average user trusts the Internet and its services. We conclude this from the fact stated previously concerning users not reading policies, while still utilising the services as intended by the companies providing them. Most people think good of others, and this apparently complies to Internet users, as well, who agree to information collection every day. With little knowledge concerning the topic, users are potentially giving away PII all over the Internet. We, therefore, recommend reading, at least parts, of provided privacy policies to get an inkling of what different companies are doing concerning the gathered information and its use.

## 7.5 Limitations

This last section includes some limitations regarding our studies and the methods used in the thesis. We encountered a couple of challenges while executing our testing, though most occurred in conjunction with the user survey.

Regarding the tools used for the testing, we experienced some variations in the results, especially when re-entering websites. This may be because the tools are not completely reliable in terms of detecting the different variations of the tracking mechanisms. Alternatively, it could be because different cookies and advertising are being set each time, for example, affected by the outcome of real-time bidding as mentioned in Section 2.3.

Our goal with performing a quantitative study was to gain enough responses that we could conclude with a generalisation for the Norwegian population. With the available time, we received just over five hundred responses. This is about half of what is needed to draw a total conclusion on behalf of Norwegian social media users. We still believe, however, that the study provides a valuable mapping of the level of knowledge for an average Norwegian user.

Lastly, we found that constructing a good user survey was slightly challenging. We focused on providing a short survey consisting of questions and alternative answers. We received feedback that stated that the respondents felt enlightened about the topic of privacy during the survey. They answered according to the alternatives presented, consequently, not answering as they might have done if they had the opportunity to provide their own answers. We still find that the survey contributed to both awareness and coaching. Though this effect may not be measurable, we find it to be a positive addition to the study.

# Chapter 8
# Concluding Remarks and Further Work

Usage of social media has gone through a change during the last years. What used to be a simplistic channel for sharing interests and ideas has become a huge platform with endless opportunities for both individuals and businesses. The basics still apply, but social media networks now offer companies opportunities to connect with individual customers, also with regards to advertising. Resulting in personalization and individualism being the drivers of the networks, and this has been made possible by the collection and processing of huge amounts of data.

The thesis has investigated how SNSs protect the users' information and if it is done in a similar fashion. Firstly, we analysed the presence of third-party companies on popular websites before moving on to test users' knowledge concerning the discussed topics. By combining this, along with the information from the background study, our goal has been to reach a conclusion on whether or not users really care about privacy on social media.

Our studies show that SNSs treat and process user information in the same fashion and by similar technologies. However, the policies do not include detailed information about how the collected data is protected by the means of encryption or other technologies. We find the differences between the SNSs in the amount of Mandatory Service Data required, and the types of data users may add to the services. Regarding voluntary information, users may provide just about anything, as long as the content is in accordance with the respective SNS policies. Hence, much of the responsibility concerning privacy lies with the users.

A large number of international companies, in addition to some Norwegian companies, collect information about Norwegian SNS users. Virtually all websites use technologies such as cookies and widgets. Consequently, information about users is constantly collected and gathered for reasons such as service enhancement and tailored advertising. Although it may not be PII that is collected, maintaining an adequate level of privacy online is difficult.

The respondents to the user survey understand the basics regarding the mechanisms used and the reasons for why information is collected online. Though, the level of knowledge concerning the extent of information sharing varies. Users claim to care about privacy and believe it is an important topic. This does not stop them from using the different services, without understanding or reading documents concerning privacy or guidelines for use of information.

Behaviour of users online shows that they do not care about privacy. They are not interested in reading documentation concerning the topic, and they do not have a clear understanding of the extent of how technologies online work. Still, several of the survey's respondents claim to care about privacy, as the majority consider the type of information they share and the visibility settings concerning this. As actions speak louder than words, we, therefore, observe that the privacy paradox, as described in Section 2.4, holds true for the findings in this thesis.

## 8.1 Further Work

As further work, we identify and propose the following;

- **Investigation of educational methods**: Our findings show that reading privacy policies is time-consuming, which has also been proven by the Norwegian Consumer Council. Additionally, we find that they provide vague information, and as a result, people rarely read them. A topic for further work could be to look into the possibilities of changing the different procedures considering privacy policies and educate people on the discussed topics. People need to gain a better understanding of the use of personal information.
- **Deeper analysis of third-party information sharing**: The analysis of third-parties in this thesis focused on the number of companies present on popular websites and did not concentrate on the type of information shared. We, therefore, believe a more detailed analysis in this area could provide highly interesting data. A potential topic for further work could be to investigate which types of user information is shared between third-parties and if it is derived directly or indirectly from users' SNSs accounts.

# References


[1] Datatilsynet. Det store datakappløpet. https://www.datatilsynet.no/globalassets/global/04_planer_rapporter/kommersialiseringsrapport.pdf. Last accessed: 2016-05-18.

[2] Wikipedia. Edward snowden. https://en.wikipedia.org/wiki/Edward_Snowden. Last accessed: 2016-05-16.

[3] Josh Kirschner. How to stop ads in your facebook news feed. http://www.techlicious.com/how-to/how-to-stop-sponsored-posts-in-your-facebook-news-feed/. Last accessed: 2016-05-30.

[4] Ipsos. Ipsos' tracker om sosiale medier q4'15. http://ipsos-mmi.no/some-tracker. Last accessed: 2016-04-06.

[5] Graig Smithl. By the numbers: 200+ amazing facebook statistics (april 2016). http://expandedramblings.com/index.php/by-the-numbers-17-amazing-facebook-stats/. Last accessed: 2016-05-13.

[6] Worldometers. Countries in the world by population (2016). http://www.worldometers.info/world-population/population-by-country/. Last accessed: 2016-05-13.

[7] Tim Grahl. The 6 types of social media. http://timgrahl.com/the-6-types-of-social-media/. Last accessed: 2016-05-12.

[8] Datatilsynet. Personvern - tilstand og trender. https://www.datatilsynet.no/globalassets/global/04_planer_rapporter/personvernrapporten-2016.pdf. Last accessed: 2016-05-18.

[9] Bruce Schneier. A taxonomy of social networking data. https://www.schneier.com/essays/archives/2010/07/a_taxonomy_of_social.html. Last accessed: 2016-05-12.

[10] Christian Richthammer et al. Taxonomy of social network data types. *EURASIP Journal on Information Security*, 2014.

[11] Wikipedia. Http cookie. https://en.wikipedia.org/wiki/HTTP_cookie. Last accessed: 2016-05-21.

[12] Vangie Beal. All about widgets. http://www.webopedia.com/DidYouKnow/Hardware_Software/widgets.asp. Last accessed: 2016-05-16.

[13] Wikipedia. Web widget. https://en.wikipedia.org/wiki/Web_widget. Last accessed: 2016-05-21.

[14] W3Schools. Html5 local storage. http://www.w3schools.com/html/html5_webstorage.asp. Last accessed: 2016-05-21.

[15] Future of Privacy Forum. All about do not track. https://allaboutdnt.com/. Last accessed: 2016-05-21.

[16] allaboutcookies.org. What is an opt-out cookie? http://www.allaboutcookies.org/manage-cookies/opt-out-cookies.html. Last accessed: 2016-05-21.

[17] BigCommerce. What is a cookie and why is it important? https://www.bigcommerce.com/ecommerce-answers/what-cookie-and-why-it-important/. Last accessed: 2016-05-21.

[18] Oxford English Dictionary. Oxford english dictionary. Last accessed: 2016-05-13.

[19] Datatilsynet. Hva er en personopplysning? https://www.datatilsynet.no/personvern/personopplysninger/. Last accessed: 2016-05-13.

[20] Data Protection Commissioner. Eu directive 95/46/ec - the data protection directive. https://www.dataprotection.ie/docs/EU-Directive-95-46-EC-Chapter-1/92.htm. Last accessed: 2016-05-13.

[21] Patricia A Norberg, Daniel R Horne, and David A Horne. The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs*, 2007.

[22] Helen Nissenbaum. *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press, 2009.

[23] Datatilsynet. Personal data act. https://www.datatilsynet.no/English/Regulations/Personal-Data-Act-/. Last accessed: 2016-04-14.

[24] Datatilsynet. Safe harbor - prinsipper om overføring av opplysninger til usa. https://www.datatilsynet.no/Regelverk/Internasjonalt/Overfoering/Safe-Harbor-prinsippene/. Last accessed: 2016-04-07.

[25] Datatilsynet. Safe harbor-beslutningen kjent ugyldig. https://www.datatilsynet.no/Nyheter/2015/Safe-Harbor-beslutningen-kjent-ugyldig/. Last accessed: 2016-04-07.

[26] European Commission. Eu commission and united states agree on new framework for transatlantic data flows: Eu-us privacy shield. http://europa.eu/rapid/press-release_IP-16-216_en.htm. Last accessed: 2016-04-14.

[27] European Parliament. Q and a: new eu rules on data protection put the citizen back in the driving seat. http://www.europarl.europa.eu/news/en/news-room/20160413BKG22980/QA-new-EU-rules-on-data-protection-put-the-citizen-back-in-the-driving-seat. Last accessed: 2016-05-30.

[28] Glenn A. Bowen. Document analysis as a qualitative research method. www.emeraldinsight.com/doi/pdfplus/10.3316/QRJ0902027. Last accessed: 2016-06-10.

[29] Inc. Alexa Internet. Top sites in norway. http://www.alexa.com/topsites/countries/NO. Last accessed: 2016-05-16.

[30] D. Crowther and G. Lancaster. Research methods: A concise introduction to research in management and business consultancy, 2009. Last accessed: 2016-06-07.

[31] Aaron Smith. Half of online americans don't know what a privacy policy is. http://www.pewresearch.org/fact-tank/2014/12/04/half-of-americans-dont-know-what-a-privacy-policy-is/. Last accessed: 2016-05-27.

[32] Øyvind H. Kaldestad. 250,000 words of app terms and conditions. http://www.forbrukerradet.no/side/250000-words-of-app-terms-and-conditions/. Last accessed: 2016-06-25.

[33] Wikipedia. Facebook. https://en.wikipedia.org/wiki/Facebook. Last accessed: 2016-04-12.

[34] Christine Jensen. Facebook-feber. http://www.aftenposten.no/digital_old/nyheter/Facebook-feber-6463588.htmlk. Last accessed: 2016-04-12.

[35] Facebook. Facebook. https://www.facebook.com/facebook/info/?tab=page_info. Last accessed: 2016-04-17.

[36] Jitender Miglani. How facebook makes money? http://revenuesandprofits.com/how-facebook-makes-money/. Last accessed: 2016-04-17.

[37] Facebook. Data policy. https://www.facebook.com/policy.php. Last accessed: 2016-04-07.

[38] Facebook. About advertising on facebook. https://www.facebook.com/about/ads/#568137493302217. Last accessed: 2016-04-08.

[39] Facebook. Social plugins faqs. https://developers.facebook.com/docs/plugins/faqs. Last accessed: 2016-04-15.

[40] Facebook. Social plugins. https://developers.facebook.com/docs/plugins. Last accessed: 2016-04-11.

[41] Rainey Reitman. How to opt out of receiving facebook ads based on your real-life shopping activity. https://www.eff.org/deeplinks/2013/02/howto-opt-out-databrokers-showing-your-targeted-advertisements-facebook. Last accessed: 2016-04-08.

[42] Facebook. Statement of rights and responsibilities. https://www.facebook.com/terms. Last accessed: 2016-04-07.

[43] Information Commissioner's Office. Key definitions of the data protection act. https://ico.org.uk/for-organisations/guide-to-data-protection/key-definitions/. Last accessed: 2016-04-11.

[44] Corporate Officer: Erin Egan. Facebook: Safe harbor certificate. http://safeharbor.export.gov/companyinfo.aspx?id=28012. Last accessed: 2016-04-17.

[45] Facebook Help Center. How do i permanently delete my account? https://www.facebook.com/help/224562897555674. Last accessed: 2016-04-11.

[46] Google. User content and conduct policy. https://www.google.com/intl/en-US/+/policy/content.html. Last accessed: 2016-04-18.

[47] Fergal Gallagher. How many users does google+ really have? http://www.techtimes.com/articles/51205/20150506/many-users-google-really.htm. Last accessed: 2016-04-12.

[48] Heather Leonard. Google+ may have a better revenue model than facebook. http://www.businessinsider.com/google-is-nipping-at-facebooks-heels-2013-1?IR=T. Last accessed: 2016-04-18.

[49] Google. Privacy policy. https://www.google.com/intl/en/policies/privacy/. Last accessed: 2016-04-19.

[50] Google. How shared endorements work. https://support.google.com/plus/answer/3403513?hl=en. Last accessed: 2016-04-18.

[51] Google. Combine personal information. https://www.google.com/intl/en/policies/privacy/example/combine-personal-information.html. Last accessed: 2016-04-20.

[52] Google. How google uses cookies. https://www.google.com/intl/en/policies/technologies/cookies/. Last accessed: 2016-04-18.

[53] Google. Google+ platform for web. https://developers.google.com/+/web/. Last accessed: 2016-04-18.

[54] Google. Buttons policy. https://developers.google.com/+/web/buttons-policy. Last accessed: 2016-04-19.

[55] Datatilsynet. Processing of sensitive personal data in a cloud solution. https://www.datatilsynet.dk/english/processing-of-sensitive-personal-data-in-a-cloud-solution/. Last accessed: 2016-04-12.

[56] Google. Self regulatory frameworks. https://www.google.com/intl/en/policies/privacy/frameworks/. Last accessed: 2016-04-07.

[57] Google. What data does google collect? https://privacy.google.com/intl/en_ALL/data-we-collect.html. Last accessed: 2016-04-11.

[58] Google. Delete your google+ profile. http://plus.google.com/downgrade. Last accessed: 2016-04-11.

[59] Google. What's happening to google+ photos? https://support.google.com/plus/answer/6262471?hl=en. Last accessed: 2016-05-04.

[60] Google. Data processing amendment to google apps agreement. https://www.google.com/work/apps/terms/dpa_terms.html. Last accessed: 2016-04-12.

[61] LinkedIn Corporation. About us. https://www.linkedin.com/about-us. Last accessed: 2016-04-18.

[62] Statista. Numbers of linkedin members from 1st quarter 2009 to 4th quarter 2015 (in millions). http://www.statista.com/statistics/274050/quarterly-numbers-of-linkedin-members/. Last accessed: 2016-04-13.

[63] LinkedIn Help. Your network and degrees of connection. . Last accessed: 2016-04-18.

[64] LinkedIn Newsroom. About linkedin. https://press.linkedin.com/about-linkedin. Last accessed: 2016-04-18.

[65] LinkedIn Corporation. Linkedin free and upgraded premium accounts. https://www.linkedin.com/help/linkedin/topics/6156/6157/71. Last accessed: 2016-04-06.

[66] LinkedIn Corporation. User agreement. https://www.linkedin.com/legal/user-agreement. Last accessed: 2016-04-07.

[67] LinkedIn Corporation. Your privacy matters. https://www.facebook.com/policy.php. Last accessed: 2016-04-07.

[68] LinkedIn Corporation. Pluginst. https://developer.linkedin.com/plugins. Last accessed: 2016-04-20.

[69] LinkedIn Corporation. Cookies on the linkedin site. https://www.linkedin.com/legal/cookie-policy. Last accessed: 2016-04-20.

[70] Yevgeniy Sverlik. Linkedin switches to custom data center design. http://www.datacenterknowledge.com/archives/2015/11/30/linkedin-data-center-in-oregon-first-to-use-next-gen-design/. Last accessed: 2016-04-11.

[71] LinkedIn Corporation. Eu data transfers and the safe harbor. https://www.linkedin.com/help/linkedin/answer/62533?lang=en. Last accessed: 2016-04-07.

[72] Statista. Number of monthly active twitter users worldwide from 1st quarter 2010 to 4th quarter 2015 (in millions). http://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/. Last accessed: 2016-04-13.

[73] Inc. Twitter. About the company. https://about.twitter.com/company. Last accessed: 2016-04-15.

[74] Twitter. Twitter privacy policy. https://twitter.com/privacy. Last accessed: 2016-04-06.

[75] Inc. Twitter. Twitter, our services, and corporate affiliates. https://support.twitter.com/articles/20172501. Last accessed: 2016-04-07.

[76] Inc. Twitter. Twitter's use of cookies and similar technologies. https://support.twitter.com/articles/20170514. Last accessed: 2016-04-15.

[77] Corporate officer: Vijaya Gadde. Twitter: Safe harbor certificate. http://safeharbor.export.gov/companyinfo.aspx?id=28200. Last accessed: 2016-04-18.

[78] Inc. Twitter. Transparency report. https://transparency.twitter.com/. Last accessed: 2016-05-06.

[79] Google. Create or change your google+ profile name. https://support.google.com/plus/answer/1228271?hl=en. Last accessed: 2016-04-14.

[80] Inc. Ghostery. FAQ. https://www.ghostery.com/support/faq/. Last accessed: 2016-04-12.

[81] Electronic Frontier Foundation. About eff. https://www.eff.org/about. Last accessed: 2016-04-11.

[82] Electronic Frontier Foundation. Privacy badger faq. https://www.eff.org/privacybadger. Last accessed: 2016-04-11.

[83] Wikipedia. Web analytics. https://en.wikipedia.org/wiki/Web_analytics. Last accessed: 2016-05-09.

[84] Analytics Market. How google analytics works. http://www.analyticsmarket.com/blog/how-google-analytics-works. Last accessed: 2016-05-08.

[85] Google. Tracking code overview. https://developers.google.com/analytics/resources/concepts/gaConceptsTrackingOverview#howAnalyticsGetsData. Last accessed: 2016-05-08.

[86] Google. Google apps (free) agreement. http://www.google.com/apps/intl/en-GB/terms/standard_terms.html. Last accessed: 2016-04-12.

[87] TNS Gallup AS. Om bruk av cookies. http://www.tns-gallup.no/om-bruk-av-cookies. Last accessed: 2016-05-08.

[88] Number 42 AS. Linkpulse. www.linkpulse.com/about/. Last accessed: 2016-05-08.

[89] Number 42 AS. Privacy policy for linkpulse. http://www.linkpulse.com/privacy-policy/. Last accessed: 2016-05-08.

[90] New Relic Inc. Let's build beautiful software. https://newrelic.com/about. Last accessed: 2016-05-08.

[91] New Relic Inc. Security for new relic browser. https://docs.newrelic.com/docs/browser/new-relic-browser/performance-quality/security-new-relic-browser. Last accessed: 2016-05-09.

[92] New Relic Inc. Privacy policy. https://newrelic.com/privacy. Last accessed: 2016-05-09.

[93] Wikipedia. Doubleclick. https://en.wikipedia.org/wiki/DoubleClick. Last accessed: 2016-05-09.

[94] Google. Doubleclick cookies. https://support.google.com/adsense/answer/2839090?hl=en. Last accessed: 2016-05-09.

[95] Joanna Geary. Doubleclick (google): What is it and what does it do? https://www.theguardian.com/technology/2012/apr/23/doubleclick-tracking-trackers-cookies-web-monitoring. Last accessed: 2016-05-09.

[96] Adform. Privacy policy and opt-out. http://site.adform.com/privacy-policy/en/. Last accessed: 2016-05-09.

[97] AppNexus Inc. Platform privacy. https://www.appnexus.com/en/company/platform-privacy-policy. Last accessed: 2016-05-09.

[98] AudienceScience. Privacy policy. http://www.audiencescience.com/privacy/. Last accessed: 2016-05-09.

# Appendix A
# Information Sheet

In this section we present the information sheet provided to the recipients before taking the user survey. The sheet is firstly presented in Norwegian, as was the original lanuage, and then an English translation is provided.

## A.1 Norwegian (Original Language)

### Personvern på Sosiale Medier - bryr virkelig brukerne seg?

#### Informasjon

Kort versjon:

- Spørreundersøkelse i forbindelse med masteroppgave på NTNU.
- Formål: Å kartlegge generell kunnskap om personvern og deling av informasjon på sosiale medier.
- Deltakelse er frivillig og all informasjon behandles konfidensielt.

Lang versjon:

Vi sender ut denne spørreundersøkelsen som en del av vår masteroppgave ved Kommunikasjonsteknologi, NTNU i Trondheim. Masteroppgaven skal gjennom tre forskningsspørsmål undersøke hva slags informasjon som samles inn av ulike sosiale medier, analysere hvordan denne informasjonen deles med andre parter, samt undersøke hva brukere av sosiale medier vet om dette.

Formålet med undersøkelsen er å kartlegge generell kunnskap om personvern og deling av informasjon på sosiale medier. Spørsmålene vil omhandle din bruk av sosiale medier, din kjennskap til vanlige sporingsmetoder og din holdning til deling av informasjon på sosiale medier.

Deltakelse i undersøkelsen er frivillig og kan avsluttes når som helst. All informasjon vil bli behandlet konfidensielt og enkeltpersoner vil ikke kunne gjenkjennes i det endelige prosjektet. Deltakere under 18 anbefales å rådføre seg med foresatte før besvarelse (foresatte kan på forespørsel få se spørreskjemaet).

Prosjektet skal etter planen avsluttes 13.juni 2016 og innsamlet informasjon vil bli slettet senest innen årsslutt.

Dersom du har spørsmål til studien, ta kontakt med Hannah (hannaher@stud.ntnu.no) eller Sølvi (solvisve@stud.ntnu.no), eventuelt veileder for prosjektet, Maria Bartnes (maria.bartnes@sintef.no)

Takk for at du deltar!

Med vennlig hilsen
Hannah Ersdal og Sølvi S. Skjærstad

Prosjektet er meldt til Personvernombudet for forskning, Norsk samfunnsvitenskapelig datatjeneste AS.

## A.2 English Translation

**Information**

Short version:

- Survey in association with a Master thesis at NTNU
- Purpose: To map general knowledge about privacy and sharing of information on social media.
- Participation is voluntary and all information will be treated confidentially.

Long version:

We are sending out this survey as part of our master thesis of Communication Technology, NTNU in Trondheim. The thesis will through three research questions examine what information is collected by various social media, analyse how this information is shared with other parties, as well as examine what social media users know about this topic.

The objective of the survey is to map general knowledge about privacy and sharing of information on social media. The questions will focus on your use of social media,

your knowledge of common tracking methods and your attitude to sharing information on social media.

Participation in the survey is voluntary and may be discontinued at any time. All information will be treated confidentially and individuals will not be recognizable in the final project. Participants under the age of 18 are recommended to consult with their parents before answering (parents may request to see the questionnaire).

The project is planned to end on June 13th, 2016, and collected information will be deleted within the end of the year.

If you have any questions, please contact Hannah (hannaher@stud.ntnu.no) or Sølvi (solvisve@stud.ntnu.no), optionally the supervisor for the project, Maria Bartnes (maria.bartnes@sintef.no).

Thank you for your participation!

Best regards
Hannah Ersdal and Sølvi S. Skjærstad

The project is reported to and has been approved by the Norwegian Centre for Research Data (NSD).

# Appendix B
# Survey Questions

In this section we present the questions from the survey. The questions are firstly presented in Norwegian, as was the original lanuage, and then an English translation is provided.

## B.1 Norwegian (Original Language)

### Personvern på Sosiale Medier - bryr virkelig brukerne seg?

#### Generell Informasjon

- Kjønn
  - Mann
  - Kvinne
- Alder
  - 13-18
  - 19-24
  - 25-29
  - 30-39
  - 40-59
  - 60-79
  - 80+
- Utdanningsnivå
  Hva studerer du nå? Eventuelt høyeste fullførte utdanning.
  - Grunnskole

- Videregående
- 3-åring bachelor
- Utdanning på masternivå/høyere grads profesjonsstudier

**Bruk av Sosiale Medier**

- Hvilke av disse sosiale mediene er du medlem av?
  Velg 1 til 4 svaralternativer.
  - Facebook
  - Google+
  - LinkedIn
  - Twitter
- Hvor aktiv er du på, dvs hvor ofte er du innom, de sosiale mediene?
  - Flere ganger om dagen
  - Èn gang om dagen
  - Et par ganger i uken
  - Et par ganger i måneden
  - Sjeldnere
  - Aldri/bruker ikke tjenesten
- Hvor ofte deler/poster du informasjon på sosiale medier (i form av tekst, bilder, o.l.)?
  - Hver dag
  - Flere ganger i uken
  - Kun ved spesielle anledninger
  - Sjeldent
  - Aldri/bruker ikke tjenesten
- Sjekker du synlighetsinstillingene på noe du deler på sosiale medier?
  Dvs. innstillingene som bestemmer om innhold deles offentlig, til venner, grupper, o.l.
  - Ja, jeg sjekker alltid
  - Jeg vet om de forskjellige innstillingene og har endret de
  - Jeg vet om de forskjellige innstillingene, men jeg har ikke endret de
  - Jeg vet ikke hvordan man endrer innstillingene

- Nei, jeg bryr meg ikke

– Er du bevisst på hvor mye personlig informasjon (navn, alder, epost, o.l.) du legger inn på sosiale medier?

- Ja, jeg legger kun inn det som er nødvendig
- Ja, men jeg legger inn det jeg vil
- Nei
- Annet, vennligst spesifiser:

– Har du personlig informasjon synlig for andre på sosiale medier?

- Informasjonen min er synlig for alle
- Deler av informasjonen min er synlig for alle
- Jeg vet ikke
- Jeg deler ikke slik informasjon på internett
- Annet, vennligst spesifiser:

**Sporingsmetoder og Deling av Informasjon**

– Utsagn: "Hvis en nettside har en "Privacy Policy" (Personvernerklæring) betyr det at de beskytter dataene dine."

- Sant
- Usant

– Har du noen gang lest gjennom en Privacy Policy?

- Ja, grundig
- Har skummet gjennom
- Nei

– Sosiale medier bruker forskjellige mekanismer til å samle inn informasjon om brukerne sine for og så selge dette videre til tredjeparter. Hvorfor?
En tredjepart vil si noen andre enn det nettstedet du kommuniserer direkte med.
Velg 1 til 5 svaralternativer.

- For å utveksle informasjon, dvs. få tilsvarende informasjon fra tredjeparten
- For å tjene penger
- For å vise målrettet reklame
- For å tilby bedre tjenester

- For å gi tredjeparten mest mulig personlig informasjon
- Ingen av alternativene/Jeg vet ikke

– Hvilke av disse nettsidene tror du Facebook samler informasjon fra?
Velg maksimum 22.

- facebook.com
- twitter.com
- linkedin.com
- plus.google.com
- vg.no
- nrk.no
- dagbladet.no
- tv2.no
- aftenposten.no
- nettavisen.no
- finn.no
- dnb.no
- yr.no
- startsiden.no
- difi.no
- sparebank1.no
- youtube.com
- msn.com
- wikipedia.org
- netflix.com
- reddit.com
- ebay.com

– De fleste nettsider benytter seg av informasjonskapsler (cookies). Hva tror du de brukes til?
Velg 1 til 4 svaralternativer.

- Cookies brukes til å kryptere pakker med informasjon
- Cookies brukes til å kartlegge en brukers aktivitet
- Cookies brukes til å lagre brukernavn og passord

- Cookies brukes til å gi bedre internettforbindelse
- Ingen av alternativene/Jeg vet ikke

– Hvor mange tredjeparter tror du gjennomsnittlig følger med på ("tracker") hva du gjør på vg.no?

- 0-5
- 6-19
- 20-29
- 30-49
- 50-69
- 70-89
- 90+

– Hvilken informasjon tror du tredjepartene samler inn, for eksempel, på vg.no? Velg 1 til 4 svaralternativer.

- Personlig informasjon
- Informasjon om din adferd
- Informasjon om brukerutstyr (type nettleser, mobil/pc, o.l.)
- Informasjon om din lokasjon
- Ingen av alternativene/Jeg vet ikke

## B.2 English Translation

### Privacy on Social Media - Do users really care?

#### General Information

– Gender

- Man
- Woman

– Age

- 13-18
- 19-24
- 25-29
- 30-39

- 40-59
- 60-79
- 80+

– Education
What are you studying now? Optionally, highest level of completed education.

- Primary School
- High School
- Bachelor's degree
- Master's degree or higher

**Usage of Social Media**

– Which of these social media sites are you a member of?
Choose 1 to 4 alternatives.

- Facebook
- Google+
- LinkedIn
- Twitter

– How active are you, i.e. how often do you visit, the different social media sites?

- Several times a day
- Once a day
- A couple of times a week
- A couple of times a month
- Rarely (More rarely?)
- Never/Do not use the service

– How often do you share/post information on social media sites (in the form of text, pictures, etc.)?

- Everyday
- Several times a week
- Only on special occasions
- Rarely
- Never/Do not use the service

- Do you check the visibility setting when sharing content on social media sites? I.e., the settings that determine whether content is shared publicly, to friends, groups, etc.
  - Yes, I always check the settings
  - I know about the different settings and I have changed them
  - I know about the different settings, but I have not changed them
  - I do not know how to change the settings
  - No, I do not care
- Are you aware of how much personal information (name, age, email, etc.) you provide to social media sites?
  - Yes, I only provide what is necessary
  - Yes, I provide the information I want to
  - No
  - Other, please specify:
- Is your personal information visible for others on social media sites?
  - My information is visible to everybody
  - Parts of my information is visible to everybody
  - I do not know
  - I do not share this kind of information on the Internet
  - Other, please specify:

**Tracking Mechanisms and Sharing of Information**

- Statement: "If a website has a Privacy Policy they protect your data"
  - True
  - False
- Have you ever read through a Privacy Policy?
  - Yes, thoroughly
  - I have skimmed through
  - No

- Social media sites use different mechanisms to collect information about their users before selling this information to third-parties. Why?
  A third-party is someone other than the website you are communicating directly with.
  Choose 1 to 5 alternatives.
  - To exchange information, i.e., receive similar information from the third-party
  - To make money
  - To show tailored advertising
  - To provide better services
  - To give the third-party as much personal information as possible
  - None of the alternatives/I do not know
- From which of these websites do you think Facebook collects information?
  Choose maximum 22.
  - facebook.com
  - twitter.com
  - linkedin.com
  - plus.google.com
  - vg.no
  - nrk.no
  - dagbladet.no
  - tv2.no
  - aftenposten.no
  - nettavisen.no
  - finn.no
  - dnb.no
  - yr.no
  - startsiden.no
  - difi.no
  - sparebank1.no
  - youtube.com
  - msn.com
  - wikipedia.org
  - netflix.com

- reddit.com
- ebay.com

– Most websites exploit cookies. What do you think they are used for?
Choose 1 to 4 alternatives.

- Cookies are used to encrypt information packets
- Cookies are used to map a user's activity
- Cookies are used to store username and password
- Cookies are used to give better Internet connection
- None of the alternatives/I do not know

– How many third-parties do you think, on average, track what you are doing on vg.no?

- 0-5
- 6-19
- 20-29
- 30-49
- 50-69
- 70-89
- 90+

– What information do you think third-parties collect on, for example, vg.no?
Choose 1 to 4 alternatives.

- Personal information
- Information about your behaviour
- Information about user equipment (browser type, mobile/pc, etc.)
- Information about your location
- None of the alternatives/I do not know