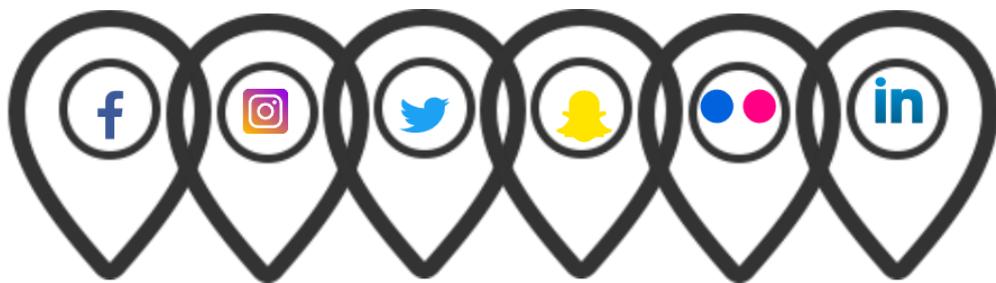


GIMA

Geographical Information Management and Applications

Privacy Paradox on Geotagging

Perception on location privacy by social media users in the Netherlands



GIMA Master Thesis
Bsc Deniz Leyla Kilic

Supervisors

Dr.Ir. Bastiaan van Loenen
Dr.Ir.Arend Ligtenberg

Professor

Prof.dr.ir. P.J.M van Oosterom

GIMA Master Thesis

Date	June 2017
Supervisors	Dr. Ir Bastiaan van Loenen Dr. Ir. Arend Ligtenberg
Professor	Prof.dr.ir. P.J.M. van Oosterom
University	University Utrecht University Utwente WUR Wageningen TUDelft
Student	Deniz Leyla Kilic
Student number	UU: 3667731 Utwente: s6026192
Email addresses:	Deniz.leyla.kilic@gmail.com Cimcime.dnz@gmail.com

Contents

Acknowledgement.....	5
Abstract	6
Chapter 1 Introduction	8
1.1 Context	8
1.2 Problem statement.....	9
1.3 Thesis structure	10
Chapter 2 Research Objectives.....	11
2.1 Research Objectives & Questions.....	11
2.2 Societal and scientific relevance.....	13
2.3 Scope	13
Chapter 3 Methodology	14
3.1 Hypothesis	14
3.2 Quantitative method: Online survey	14
3.3 Operationalization	15
Chapter 4 Location information as personal data	18
4.1 Location information	18
4.2 Definition and conceptualization of privacy.....	18
4.3 Location information as personal data.....	20
Chapter 5 Geotagging functionality in social media.....	22
5.1 Geotagging	22
5.2 Technology behind geotagging.....	25
Chapter 6 Monetization of privacy and its risks	27
6.1 Monetization of personal data	27
6.2 Collection of personal data and location data.....	28
6.3 Dimensions of location privacy problems on social media	30
Chapter 7 Privacy paradox: behaviour, attitude and concerns	33
7.1 Behaviour.....	33
7.2 Concerns.....	35
7.3 Attitude.....	35
Chapter 8 Sample	39
8.1 Characteristics of the participants	39
8.2 Response rate	39
8.3 Representation analysis.....	41

Chapter 9 Results.....	44
9.1 Geotagging behaviour on social media	44
9.1.1 Geotagging	45
9.1.2 User characteristics	47
9.1.3 Places.....	49
9.1.4 Motivations.....	50
9.2 Attitude.....	52
9.2.1 Audience of location information.....	52
9.2.2 Re-use of location data by third parties	53
9.2.3 Privacy attitudes: interest positions	54
9.2.4 Dichotomy between privacy interest position and willingness to geotag.....	55
9.3 Concerns.....	55
9.3.1 Concerns regarding location privacy	55
9.3.2 Dichotomy between concerns and willingness to geotag	56
Chapter 10 Conclusion	57
Chapter 11 Discussion	61
11.1 Reflection.....	61
11.2 Recommendation	62
References	64
Appendices	70
Appendix 1 Online survey.....	71
Appendix 2 Response rate and missing values.....	78
Appendix 3 Analysis Scheme	81
Appendix 4 Likert Scale.....	85
Appendix 5 Item analysis for multiple response questions.....	92

Acknowledgement

This research is established with the help of number of people I would like to thank. I would first to thank my supervisor Bastiaan van Loenen for providing his guidance and feedback. Our meetings generated new insights and interesting conversation about privacy and data protection, which keep me on track and motivated. The monetization of personal data by location-based services and the user's right to privacy gained my interest. I would like to thank my second supervisor Arend van Ligtенberg for his time and guidance throughout the course of this research, which also keep me on track. I would also like to thank Peter van Oosterom for his involvement.

I would also like to thank all the participants for their time and their contribution to this research. In three weeks, we managed to get 184 participants. This accomplishment would not have been possible without them. Finally, I must express my gratitude to my parents, Wouter van de Hoef and Feridun Kilic providing me support throughout the process of researching and writing this thesis.

Enjoy your read,
Deniz Kilic
June 2017

Abstract

Nowadays, social media applications are embedded with geotag features, and users share their experiences of places with other people online to express themselves. Users post their text messages, photos and videos online and geotag them to share the location of the undertaken activity. Geotag features enhance the user experience, but also introduce privacy related threats and issues. Social media users place themselves at a greater risk for identity theft or cyber stalking by disclosing their whereabouts online. Academic literature suggests social media users are concerned about their online and location privacy, and also have a cautious attitude towards sharing personal and location information. However, users keep sharing their location information with fellow users and application services despite their concerns. This dichotomy between concern and disclosure behaviour is known as privacy paradox.

This thesis investigates if privacy paradox applies to geotagging behaviour on social media. With the help of an online survey, 184 social media users are questioned about their geotagging activities, their attitude towards re-use of location information and the accessibility of their location information by fellow users. In addition, their concerns regarding location privacy are also examined.

This thesis suggest that the privacy paradox does not apply to geotag behaviour on social media since participants rarely geotag, and when they do, it is for special occasions such as vacations and trips. Out of 149 participants, 37,6% claimed they do not geotag due to caution for over sharing personal information or to keep themselves anonymous. 62,4% of the participants reported geotagging their content on social media by using points of interest or manually adding a location.

In general, participants displayed neutrality towards privacy threats. Therefore, there is no privacy paradox between user's geotagging behaviour and their concerns regarding location privacy. The type of place and the type of relation with fellow users determine whether others are allowed to see a participants' whereabouts on social media. The same findings also apply to participants' attitudes towards re-use of location information by parties such as advertisers, the government, and intelligence services. In general, participants are cautious regarding the accessibility of their location information to fellow users and third parties. Out of 135 participants, 58,8% claimed themselves as privacy pragmatists, 27,4% as fundamentalist and 11,1% unconcerned. The association between the self-reported privacy interest position and willingness to tag is moderately strong. This association means that there is no privacy paradox between the attitude towards location privacy and the willingness to geotag content on social media.

Users still do not have control of their own personal data flow and are not fully aware for what purposes their data is used, or which third parties have access to their personal data. Social media companies are not transparent about the monetization of personal data. To resolve this informational asymmetry between users and social media companies, privacy awareness among users should be stimulated via guidelines and tutorials with information. The guidelines should include information about the following subjects: (1) collection of location data, (2) storage of location data, and (3) third parties and re-use of location data. For further research on geotagging behaviour and privacy awareness recommendations are suggested.

Chapter 1 Introduction

1.1 Context

Online social networks, also known as social media, have become increasingly popular and gather millions of users (Scellato, Mascolo, Musolesi & Latora, 2010). Social media such as Facebook, Twitter and Instagram engage their users to produce, share and consume information over social links. The scale at which content is produced may be considered overwhelming. For example, Facebook has more than 400 million active users who share about 3.5 billion pieces of content each week, and upload 2.5 billion photos each month (Scellato et al., 2010). These online activities are also known as *micro-blogging*, which has evolved from a trend to a daily activity for users. The users of social media share their photos, videos and text messages online with other users on the internet (Tang, Lin, & Hong, 2010) for several reasons such as networking, relationship development and self-representation (Lee, Park, & Kim, 2013).

Nowadays, the majority of social media is enabled with location services due its rapidly growing popularity on the web. Positioning technologies such as GPS and WIFI triangulation have become the standard functionality for mobile devices and enables users to enjoy the ease of social media on a mobile device (Friedland & Sommer, 2010). Location-based services allow users to continuously collect and share their location with fellow mobile device users and companies (Mascetti, Freni, Bettini & Wang, 2011). The merging of location information with content online is often referred to as Geoweb¹ (Elwood & Leszczynski, 2011). Likewise, social media with location-based services may also be referred to as *location-based social networks* (Mascetti et al., 2011).

Social media platforms such as Facebook, Twitter and Flickr allow users to determine their location and share their location information with the web and fellow users. This activity is also called geotagging. This means users can assign place names, geographical coordinates or any other locational information to text, images, videos or other content on the web (Elwood & Leszczynski, 2011). Geotags may be manually ascribed or automatically generated by applications that run on mobile device. This geo-tagged data, created by the user, contains location information which provides the position of a person or attribute at a certain point in time and within certain accuracy (Longley, 2001, as cited by van Loenen, de Jong, & Zevenbergen, 2008). In terms of geoweb, this means users of social media are also producers of location information.

Social media with location services brings advantages for the experiences of users, but it has also introduced privacy related issues and threats (Mascetti et al., 2011). Acquisti and Gross (2006) (as cited by Young & Quan-Haase, 2009) stresses how social media users place themselves at a greater risk for cyber and physical stalking, identity theft and surveillance by disclosing their personal information on social media. Considering the accessibility to tremendous amounts of data and the available technologies for identifying a person or even cybercasing, publishing location data is only one part of the privacy problem and it requires necessary data protection.

According to Friedland & Sommer (2010), the threat to privacy is elevated to a new level by the combination of three related developments. Firstly, the relatively small percentage of online videos and images with location data is sufficient for mounting systematic privacy attacks. Secondly, everyone can sift through large volumes of geo-tagged data without much effort thanks to the availability of location-based search capabilities. Thirdly, the availability of so many other location-based services and annotated maps such as Google Street View, allows individuals and actors to find correlation across diverse independent sources (Friedland & Sommer, 2010), which makes citizens with a social media profile vulnerable to identity fraud, cybercasing and surveillance (Young & Quan-Haase, 2009).

¹In the broader sense, Geoweb also refers to the mobile devices as hardware, applications and services as software and programming techniques that include such as APIs and interactive mapping platforms.

However, social media users are not fully aware of, or have little knowledge about these privacy risks within disclosure of location information on social media. Users tend to have relatively high concern about their online privacy (Kokolakis, 2015). On the other hand, users reveal a lot of personal information on social media for relatively small rewards and the attention of peers (Kokolakis, 2015). Users are also willing to share their location information on mobile devices as long as the service is useful (Barkhuus & Dey, 2003). Danezis, Lewis & Anderson (2005) and Cvreck & Kumpost (2006) studied under which circumstances citizens are willing to share their location information (Cvrcek & Kumpost, 2006; Danezis et al., 2005). Both studies have common findings: users have low privacy expectations, and location information can be acquired against a relatively small monetary return. There appears to be a dichotomy on this matter. Users tend to have high privacy concerns, but this doesn't seem to affect their online behaviour. This dichotomy is also known as privacy paradox (Pavlou, 2011) and is often referred as the inconsistency of privacy concerns and disclosure behaviour.

1.2 Problem statement

Privacy paradox has been studied and interpreted through diverse theories such as privacy calculus models (Jiang et al., 2013), cognitive biases in decision-making (Cho et al., 2010), information boundary theory (Acquisiti & Gross, 2006) and structuration theory (Kokolakis, 2015; Zafeiropoulou, Millad, Webber & O'Hara al., 2013). The privacy calculus model argues that individuals perform a calculus between the expected loss of privacy and the potential gain of disclosure (Jiang et al., 2013). Their final decision is determined by the outcome of the privacy trade-off (Jiang et al., 2013). When one considers the conclusions derived from the studies of Danezis et al. (2005) and Cvreck et al. (2006) the disclosure behaviour becomes understandable. Several studies have confirmed that users of social media weigh the risks and benefits of sharing private information, and disclose information when the services and benefits outweigh the observed risks (Cho, Lee & Chung., 2010).

An important critique on the privacy calculus model can be made. The model assumes that citizens make decisions as rational agents by calculating risks and benefits. Unlike in the privacy calculus model, behavioural economics has shown that decision making by individuals is affected by cognitive biases and heuristics. In other words, individuals tend to be overconfident or have an optimistic bias. Individuals display a strong optimism bias about online privacy risks. They judge themselves to be less vulnerable than others for risks such as cybercasings, identity theft and surveillance (Cho et al., 2010). Some individuals are indifferent to the privacy risks of disclosure information. Besides, not everyone has access to all necessary information to make informed judgements about the privacy trade-off (Kokolakis, 2015). Privacy decisions are made in limited time with incomplete information about risks and benefits. In other words, the privacy decisions are constrained by bounded reality and incomplete information (Acquisti & Grossklags, 2006). This limited access to information about privacy risks may also be referred to as informational asymmetries. Informational asymmetries refer to the relationship between consumers and providers in the online and mobile market (Kokolakis, 2015). Consumers are not fully aware, or do not have knowledge of how their personal data is used and disseminated to third parties (Martijn & Tokmetzis, 2016). In addition, it is quite hard to get an overview or gain information on how personal data is stored and disseminated to other companies and organizations (Martijn & Tokmetzis, 2016). This may be one of the essential reasons why users display inconsistent behaviour towards their privacy concerns. Because of the lack of information about data flow, people are not aware of the privacy risks on social media (Nissenbaum, 2011).

The structuration theory of Giddens (1984) can be used to explain the privacy paradox (Kokolakis, 2015). The structuration theory claims human agency and social structure are dependent upon each other. Agency means the ability of humans to act on free choice, while social structure refers to contextual factors which can be stimulating or act as a constraint during the decision-making process of an individual. Decision-making is a process of structuration. This means people do not make location-

sharing decisions as an entirely free agent, but are heavily influenced by external factors such as social norms or trust in the social platforms during the privacy trade-off decision (Zafeiropoulou et al., 2013).

Geotagging and privacy paradox

Privacy paradox has mainly been investigated by studying web users' disclosure behaviour of personal information in diverse contexts such as online social networking, online shopping and location-based applications for mobile devices through diverse theories (Kokolakis, 2015). Recently, location privacy on social media has attracted more and more attention from academia and industry. However, privacy paradox on sharing location information and location privacy on social media has been given little attention (Alrayes & Abdelmoty, 2014; Freni, Ruiz Vicente, Bettini & Jensen, 2010; Mascetti et al., 2010). For this reason, location privacy should be examined more closely in a social media context.

Geotagging is an important function of location-based social networking services, where users share their locations by checking in at places to let their connections know where they are (Lane & Walton, 2008). The question remains whether the privacy paradox does exist for geotagging behaviour on social media. If this is the case, how can the inconsistency between, on the one side, the privacy concerns and attitudes of individuals, and on the other side, behaviour of individuals, be solved?

1.3 Thesis structure

Chapter 2 explains the research objectives, the main and sub questions. Chapter 3 provides the methodology of this research by explaining the steps taken in order to answer research questions and achieve these objectives. Chapters 4, 5, 6 and 7 address the theoretical sub questions and their objectives by describing location information, geotagging on social media and user privacy trade-off based on a critical account of existing academic literature. Chapter 8 provides details about the survey sample in terms of non-response rate and the characteristics of the participants who joined the research. Chapter 9 illustrates the results obtained from the data collected via online survey. Chapter 10 answers the research questions and gives an overall conclusion for the thesis. In the final Chapter, the findings of the thesis and its relation to academic literature is discussed. In addition, improvements on methodology are discussed as well. The thesis ends with a recommendation for further research on geotagging behaviour in social media and privacy issues.

Chapter 2 Research Objectives

This chapter introduces the research objectives and research questions (Section 2.1). The societal and scientific relevance of this thesis will be explained (Section 2.2) and the scope of the research will be summarised in the final section (Section 2.3).

2.1 Research Objectives & Questions

2.1.1 The objectives

This research aims to investigate whether the theory of privacy paradox applies to location information by studying users' concerns on location privacy and their geotagging behaviour on social media. With new insights, recommendations will be made on how to overcome the expected gap between the privacy concern and behaviour regarding location information. The motivation of this research is to contribute to the establishment of privacy awareness on location privacy in the social media and personal data protection.

The research consists of three parts. The first part is a literature study about location information, geotagging behaviour on social media, privacy attitude and concerns. The second part of the thesis investigates the geotagging behaviour of social media users via survey. In the third part the privacy paradox will be further explored and alternatives provided to better balance the privacy interests of individuals with their social media needs.

Three objectives are identified in this research:

- 1) Define the link between location information on social media and data protection;
- 2) Define the privacy paradox on geotagging in the context of social media.
 - a. Examine attitudes toward location information
 - b. Examine privacy concerns regarding location privacy
 - c. Study geotagging behaviour on social media
- 3) Finding new insights on the privacy paradox on location information in social media and formulating directions for addressing such paradox.

2.1.2 The research question

The main research question of this thesis is: *Does the privacy paradox theory apply to geotagging on social media, and if so how may the privacy paradox be overcome?*

The main question is divided into four sub questions:

- 1) *To what extent is location information considered to be (sensitive) personal data by law?*
 - a. What is location information?
 - b. Under which circumstances is location data considered as personal data?

The first question explores the definition of location information and its sensitiveness according to the law in the European Union through desk research and scientific literature study.

- 2) *What is geotagging?*
 - a) *How do geotag features work on social media?*
 - b) *For what reasons do social media companies collect personal data?*
 - c) *What are the risks of geotagging for the privacy of social media users?*

The second question defines geotagging in general and specifically in the context of social media. Through literature study and desktop research applications of geo-tagged data in the business area, geotagging will be explored. In this research, the users are defined as the individuals who use social media and create content on web 2.0. However, from a volunteered geographical information perspective, social media users are intentionally or unintentionally also “producers” of geotagged data. Therefore, the privacy implications for the “producers” of geo-tagged data and its applications will also be studied.

3) *What is the self-reported geotagging behaviour on Social media?*

This question will be also examined through the survey and it aims to explore self-reported geotagging behaviour on social media services such as Facebook, Twitter and Instagram. The geotagging behaviour regards the users’ motivations, the frequency of geotagging and which social media service one uses to geotag their content.

4) *What are the attitudes and concerns of users on location privacy when they are online on their social media profiles?*

The fourth question aims to explore the attitude and concerns of users towards location privacy on social media. One’s attitude and concerns may be an outcome of his or her perception on location privacy and to what extent one may consider location as personal information. The concerns of a user are also related to his or her knowledge about the risks of disclosure location information on social media. The third question will be examined through an online survey.

Table 2-1 Overview objectives and sub questions

	Objective 1: Define the link between location information on social media and data protection		Sub question 1
	Objective 2: Define the privacy paradox on geotagging in the context of social media		Sub questions 2,3 and 4
	Objective 3: Finding new insights on the privacy paradox on location information in social media and formulating directions for addressing such paradox.		Sub questions 1, 2, 3 and 4

The results of the third and fourth research questions will be analysed to investigate whether people’s geotagging behaviours are paradoxical compared to their privacy attitudes and concerns regarding location privacy. Sub question 1 corresponds to the first objective (Table 2-1). Sub questions 2, 3 and 4 correspond to second objective. The findings for the research questions correspond to the third objective.

2.2 Societal and scientific relevance

Privacy is a multidisciplinary and diverse concept. As such, it requires diversity in research on matters related to privacy. It also requires active collaboration and integration with other fields, and an inimitable understanding of cultures and regulations (Dinev, 2014). Unfortunately, privacy paradox has only been studied in isolation (Kokolakis, 2015). Meanwhile, a better understanding of the privacy paradox may enable a new perspective on the legal and ethical framework of information privacy. This may be achieved when the paradox is studied in relation to the technological environment, privacy awareness campaigns or the availability of privacy enhancing technologies (Kokolakis, 2015). Identifying a privacy paradox for location information will also call for new ways to address users' needs in using social media. Location plays an important role as a catalyst, allowing systems to infer new data and information about individuals, often without their knowledge and consent. Due to this attribute, people's attitude to their location data are important to study (Zafeiropoulou et al., 2013). Addressing the privacy paradox or minimizing the paradox may result in new niche products respecting one's privacy and offering the benefits of social media.

In addition, it is important to understand the legal framework, since in the age of globalization and global data sharing and trans-border data flow, various national regulations can pose significant problems (Dinev, 2014). These regulations will play an important role in what is considered as privacy data within the emergence of global data standards and interorganizational data exchange (Dinev, 2014). In other words, the regulations will influence the exchange of data, data synchronization and interorganizational systems (de Corbière and Rowe, 2013, as cited by Dinev, 2014)

2.3 Scope

Most scientific work has been focused on the existence of privacy paradox in online shopping, location-based service applications and online social networking (without location services). This thesis will not focus on privacy paradox in those settings. The perception of social media users on location privacy and their geotagging behaviour is the main focus of this thesis.

Chapter 3 Methodology

This chapter describes how the research is carried out. The hypothesis and the concepts of theory are operationalized (Section 3.1). The snowball sampling (Section 3.2), online survey and its design (Section 3.3) is discussed as well.

3.1 Hypothesis

The thesis contains four research questions related to the privacy paradox and geotagging behaviour on social media. The first and second questions are related to background information and theories about privacy, location information and social media. As mentioned before in the objectives (Chapter 2) a literature and desktop research are applied. The results from these studies gives directions and creates a basis for formulating the hypothesis and the online survey.

Meanwhile the third and fourth questions are empirical, and are answered with quantitative methods. In a quantitative research setting, theories are tested in a deductive way through structured methods such as surveying or content analysis (Bryman, 2012). Through quantitative studies the researcher may reach many respondents, and quantify a social structure in numbers, finding correlations between phenomena (Bryman, 2012).

To answer the empirical sub questions 3, 4, and the main research question, a hypothesis is formulated. The formulation of the hypotheses is based on the theoretical framework. The expected outcome of this study is expressed with the hypothesis, which means the hypothesis will be the basis of the survey: **The privacy paradox does exist for geotagging behaviour in social media.**

Previous studies show users of location-based social networks disclosing their whereabouts with fellow users for several reasons such as self-representation, relation development and social control, whereas the users are also concerned about their online- and location privacy references. Since geotagging of content is a way to disclose location information on the web one may expect the privacy paradox also applies to geotagging behaviour on social media.

3.2 Quantitative method: Online survey

With the help of online survey this research investigates whether the **privacy paradox** does indeed exist for **geotagging behaviour in social media**. Online surveys have the potential to access a large size sample of population. Secondly, it is more cost effective compared to other techniques (Reips, 2002, as cited by Rahman, 2016). Third, the data gained from the online survey is already digitised, which saves time. The online survey is made via Thesistools.com, which enables downloading data in the Excel format, allowing it to be imported into the statistical predictive software program SPSS.

The target group of the survey are social media users who possess a mobile device such as a smartphone or a laptop. One should consider that privacy paradox is not apparent in of young users, but concerns users of all ages (Kokolakis, 2015). For this reason, the sample should contain an as diverse as possible sample of users to detect the privacy paradox on geotagging behaviour. To reach as many social media users as possible the snowball sampling is used. Snowball sampling is defined as “*a sampling technique in which the researcher initially samples a small group of people relevant to the research questions, and these sampled participants propose other participants who had the experience or characteristics relevant to the research*” (Bryman, 2012). The sampling is not random and is based on nonprobability sampling. Nonprobability sampling is also known as purposive sampling, which means that the participants are chosen with a certain property with a specific purpose in mind (Trochim, 2017).

The snowball sampling method has advantages and disadvantages regarding research and its representativeness and quality of results. Snowball sampling is an appropriate method to reach a certain group with specific characteristics and may lead to a relatively high response rate among social media users in the Netherlands (Trochim, 2017). Due to the high response of the survey the useable

results can be utilised for further research. As a result of snowball sampling it is possible to spread the survey rapidly to large numbers of social media users online. Persons who received an invitation for the survey were additionally asked to forward the survey to friends, families and co-workers.

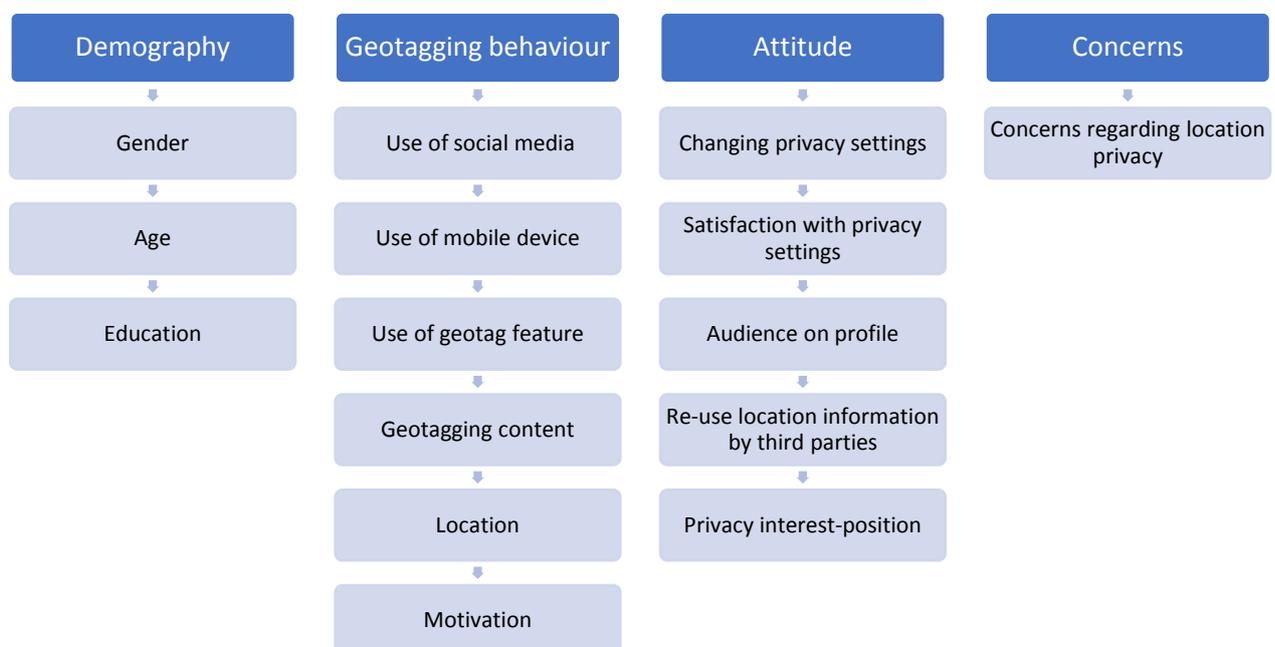
Online surveys have several limitations, particularly with the representativeness of the sample and selection biases, difficulties in measuring non-response rate and a lack of control of the testing environments. Due snowball sampling, the study population is not representative of the entire population of the Netherlands. Because of this one cannot generalize statements or draw conclusions related to the entire population (Bryman, 2012). However, it is possible to conceive suggestions based on theoretical saturation and analytical generalization for policies and management. It is also possible to generate new theories about complex subjects based on the results and their analysis. The findings do provide a basis for exploration of geotagging behaviour, attitudes, and concerns regarding location privacy.

There is also a chance for community bias to occur in the study population due to snowball sampling. The first participants of the survey may influence the sample. To reduce a community bias, the researchers must keep the information flow going throughout the target group. Snowball sampling is a method to investigate whether privacy paradox does exist on geotagging by users in social media. It is also not possible to determine the sampling rate for the survey due to lack of information regarding the number of persons who have had the opportunity to participate. Due to this it is also not possible to calculate the total non-response rate. However, it is possible to calculate the non-response rate for each question of the survey (see Chapter 8).

3.3 Operationalization

Firstly, the operationalization of the theme's related to privacy paradox are shortly explained. Secondly, the design of the questions for each theme is described (Figure 3-1).

Figure 3-1 operationalization of theme's



The hypothesis lays down the basis for designing the survey (Appendix 1). The survey has 15 questions in total (Table 3-1). The first three questions are related to demographic characteristics such as age, gender and level of education. The self-reported geotagging behaviour of the participants is investigated from questions 4 till 9. The types of question are diverse, ranging from multiple choice to

Likert-scales. Likert-scales are commonly used to measure a population's subjective interpretations, attitudes and opinions (Rahman, 2016). The questions are formulated as statements and the respondent gives an answer within a level of agreement for the given statement (Rahman, 2016).

Table 3-1 Overview questions and operationalization

Theme	Question	Operationalization	Type	Theory Chapter
Demographic	1	Gender	Multiple choice	Section 7.3
	2	Age	Open question	
	3	Education	Multiple choice	
Geotagging behaviour	4	Use of social media	Likert-scale	Sections 5.1, 7.1
	5	Use of mobile device	Likert-scale	Sections 5.1, 7.1, 7.3
	6	Use of geotag feature	Multiple choice	Section 5.1
	7	Geotagging content on social media	Likert-scale	Section 5.1, Chapter 7
	8	Location type	Multiple response	Chapter 7
	9	Motivation to geotag	Open question	Chapter 7
Attitude	10	Changing privacy settings	Multiple choice	Section 7.3
	11	Satisfaction with privacy settings	Likert-scale	Section 7.3
	12	Audience on profile	Multiple response	Section 7.3
	13	Re-use of location information	Multiple response	Section 7.3
	14	Privacy interest-position	Multiple choice	Section 7.3
Concerns	15	Concerns regarding location privacy	Likert-scale	Chapter 6 and paragraph 7.2

Geotagging behaviour

Within question 4 and 5 the participants were asked which social media and mobile device they use, and how they managed their location information on social media. Social media is defined as the mostly used social platforms such as Facebook, Twitter, Instagram, Snapchat and Flickr. The mobile device is categorized as a smartphone, tablet and laptop or computer. Both questions are Likert-scales with 5 item options: "Never", "rarely", "Sometimes", "Regularly" and "Continuously".

The geotagging behaviour is defined in two ways: use of geotag features on social media and the content that is geotagged by the participants. In question 6 the participants were asked to indicate which geotag feature they use on social media. The geotag features of Facebook are used as an example in the survey, since Facebook is still one of the most used social media networks (van der Veer, Boeke & Peters, 2017). The question has the following options: "List with suggestions", "Add manually places", "Both", "None" and "I don't know".

Questions 7 and 8 are about places and how many times participants geotag photos, videos and text. The willingness and motivation to disclose location information on a social network highly depends on the type of place and on the recipients in the network (Wagner et al., 2010). Question 7 about geotagging content is a Likert-scale with 5 item options: "Never", "Rarely", "Sometimes", "Regularly"

and “Continuously”, whereas question 8 is a multiple-choice question with suggestions of different places such as vacation, trip, home, work and hospital.

As a follow-up, the participants were asked why they geotag their content on social media in question 9, which is an open question. Users on social media share their location information for many reasons: communication and coordination purposes (Wagner et al., 2010), self-representation and promoting a certain lifestyle (Barkhuus et al., 2008). Cramer, Rost and Holmquist (2011) also discovered that users checked-in at Foursquare, a location-based application, for discounts and to discover new venues.

Privacy attitude and concerns

Question 12 touches upon attitudes of users on social media. The location type and viewer of location information is categorized in the survey due to the context-dependence of privacy. The perception of location information of an individual depends on the undertaken activity and its location such as home, work, and city centre, but also on the accessibility of this location information on the user’s profile to other followers. These followers may be family but also co-workers. The way one manages their privacy depends on his or her relation to others. Individuals have different perceptions of privacy, and their expectations may change due a given context. The theory behind question 12 is optimized with a multiple response system. The audience is divided into many different social relations such as friends, family, colleagues, employers and acquaintances. However, it is also possible one prefers that nobody can know their whereabouts. To give this as an answer participants may choose option: nobody. The suggested places are the same as in question 7: home, work, trip, hospital and political event.

Question 13 touches upon re-use of location information by third parties. Due to the context-dependence of privacy, respondents are asked to choose which location information may be seen and reused by whom and for which purposes. The suggested third parties are derived from chapter 6 about the monetization of personal data. Social media companies claim they collect personal data (including location information) to improve and personalize services with relevant advertisements and recommendation mechanisms. Advertisers use personal data for more personalized advertisements, whereas government or researchers use personal data for policymaking or studies.

Question 14, which is a multiple-choice question, comes with a short description about three kinds of privacy attitudes, namely privacy fundamentalist, pragmatist and unconcerned. The respondents are asked to answer which description of attitude matches best with her or his attitude. These three short descriptions of three persons are based on the ideological-interest positions on privacy based on Westin et al., (2003; as cited van Loenen, de Jong & Zevenbergen, 2008).

The final question, 15, is Likert-scale with 7 items. The 7 items are based on the theoretical chapter 6 and 7 about the monetization of personal data and the concerns of users regarding their privacy. These concerns are operationalized with statements about privacy threats and re-use of location data by third parties. The privacy threats are defined as “identity fraud”, “issues with social relations” and “hackers”. In addition, the re-use of location data by government, companies and research is also used as an item. The participants were asked to indicate to what extent they agree with the statements with these options: “ Strongly disagree”, “Disagree”, “Agree”, “Neutral”, “Agree” and “Strongly agree”.

Chapter 4 Location information as personal data

This chapter aims to answer the first sub question:

- 1) *To what extent is location information considered to be (sensitive) personal data by law?*
 - a) *What is location information?*
 - b) *Under which circumstances is location data considered to be personal data?*

The definition of location information and privacy will be first described according to literature (4.1 and 4.2). The reason why location data is considered personal data according to GDPR will also be explained (4.3).

4.1 Location information

The thesis follows the definition of location information as it is defined by E-Privacy Directive 2002/58/EC: *“(Article 14) Location data may refer to the latitude, longitude and altitude of the user's terminal equipment, to the direction of travel, to the level of accuracy of the location information, to the identification of the network cell in which the terminal equipment is located at a certain point in time and to the time the location information was recorded.”* Location data is processed in an electronic communications network and indicates geographic position of the terminal equipment of a user in a publicly available service (Article 2 of E-Privacy Directive 2002/58/EC).

Due to the development of communication technology such as mobile devices, it is presently possible to monitor human behaviour, attitudes and preferences on a large scale with the help of methods for mining and cataloguing of personal data. One important feature of mobile devices is their ability to collect location information. A mobile device needs to connect to telecommunication towers to be connected to the telecom infrastructure. The location of the telecommunication tower or a combination of towers results in a location of the device (the cell-ID). Moreover, a mobile device can be equipped with sensors such as a camera, a GPS receiver or Wi-Fi wireless interface to identify the user's current location. Social media utilizes web 2.0 technology to generate personal data for millions of users including location information (Chang & Sun, 2011). Social media users are also producers of their own location information when they geotag their digital content such as photos (Goodchild, 2007). In addition, they can share their location information with application services and third parties in return for "free" services or products (Acquisti, Brandimarte & Loewenstein, 2015). In other words, users are involved in a privacy trade-off where the users share their location information in return for services. To determine to what extent location information might be considered as sensitive information the definition of privacy and personal data will be explained first.

4.2 Definition and conceptualization of privacy

According to Allen (1988) it is difficult to capture the exact meaning of privacy in words because it is an elastic concept. It is possible to develop different definitions of privacy depending on one's perception on what information should be private and public (as cited by van Loenen & Zevenbergen, 2007). Due to the unclear boundaries of privacy it is hard to define the relation between privacy and other cognate concepts such as anonymity (Marglious, p.415, 2003 as cited by van Loenen & Zevenbergen, 2007). 'The right to be left alone' is one of the most well-known definitions of privacy; however, it is a broad concept. Meanwhile other academics such as Westin (2003) and Nissenbaum (2010) have sharpened the definition of privacy.

Many definitions of privacy share a common core of key elements (Marglious, 2003, p.415, as cited by van Loenen & Zevenbergen, 2007). One of those key elements is control over the transaction of personal information through interactions and communications that regulate access to this

information. When the level of control is relatively high over these transactions, the vulnerability decreases and the options for making decisions increase (Margliu, 2003, p.415, as cited by van Loenen & Zevenbergen, 2007).

Westin (2003) defined privacy as "the claim of an individual to determine what information about himself or herself should be known to others." (as cited by Cottril & Thakuria, 2012). This definition references in which context the disclosure of information should take place. Within the concept of *contextual integrity* Nissenbaum (2011) explains how privacy is context-dependent. The norms of privacy may differ depending upon who is on the receiving end of the information flow, the type of disclosed information, and the uses to which the shared information will be put. According to Nissenbaum (2011) privacy is a right to appropriate flow of personal information.

Controlling or regulating the access to information can be distinguished into four types of privacy rights (Figure 4.1):

(1) The privacy of the body refers to the protection of people's physical selves against invasive procedures such as drug administration or generic tests and cavity searches.

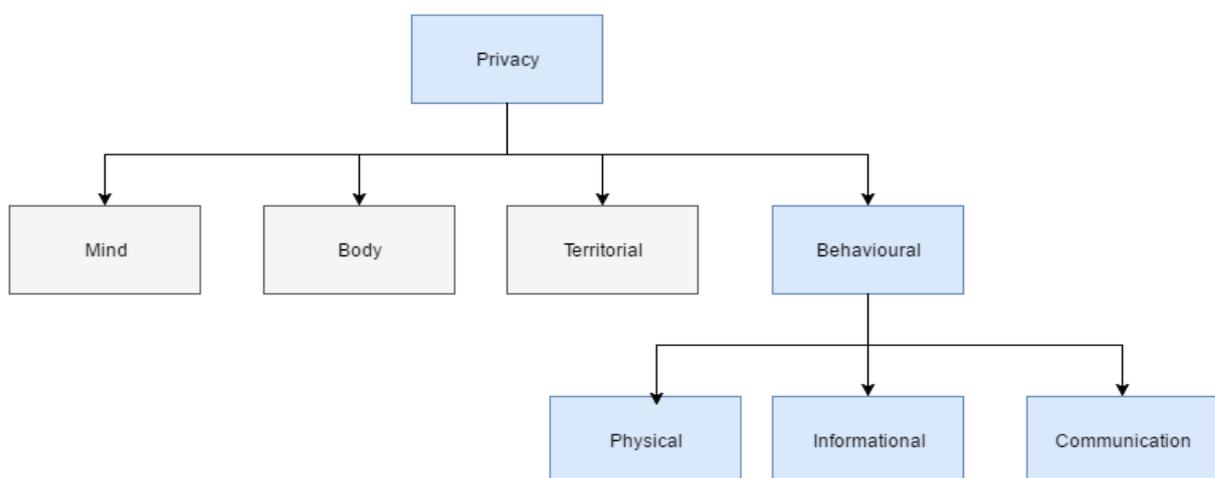
(2) The privacy of the mind/psychological state refers to the right to have freedom to think and keep information someone does not want to reveal about themselves, to themselves.

(3) The territorial privacy refers to the limited settings on intrusion into places such as home and workplace.

(4) The behavioural privacy refers to the right of one to behave as one likes. Behavioural privacy can be divided into three categories:

- Physical privacy, also known as personal privacy: privacy as a right to have freedom of movement: a state or conditions of limited physical access to a person;
- Informational privacy; privacy as a right to control access to and dissemination of information about oneself
- Privacy of communications.

Figure 4-1 Types of privacy



Source van Loenen & Zevenbergen, 2007

Smith, O'Hara and Lewis (2011) differentiate between physical and informational privacy by stating that physical privacy concerns physical access to an individual and the private space, whereas information privacy stresses out the accessibility to individually identifiable personal information and the right to control the flow of personal information. In other words, the informational privacy addresses to what extent one is able to control the use of his or her personal information (as cited by van Loenen et al., 2008).

Location privacy

Location privacy is a special type of informational privacy and is also known as geoprivacy. Individuals have the right to protect their location information from disclosure or determine the extent to which their data can be shared (Duckham and Kulik, 2006; Sila-Nowicka and Thakuriah 2016). Location privacy encapsulates the idea that an individual whose location is being tracked should control who can know it (Krum, 2008).

4.3 Location information as personal data

Informational privacy is also known as privacy of personal data (Pötzsch, 2009; van Loenen et al., 2008). Personal data is defined as *"information relating to an identified or identifiable natural person"* by the European Data Protection Directive 95/46/EC. According to the directive an identifiable person is one who can be identified in a direct or indirect way by reference to an identification number, or to one or more factors related to his or her physical, physiological, mental, economic, cultural or social identity (Article 2(a) of the EU Data Protection Directive, as cited by Loenen et al., 2008). This means that personal data is more than a name, address or telephone number, but it also may contain other elements which have economic, cultural or a social component. The physical, physiological and mental factors are for example related to one's health status.

The Data Protection Directive 95/46/EC doesn't frame location data as personal data, whereas the new legal framework General Data Protection Regulation (GDPR) in the EU does recognize location data as personal data. According to GDPR the definition of personal data is also expanded to other online data as well, such as Cookie IDs and IP addresses: *'personal data' means any information relating to an identified or identifiable natural person 'data subject'; an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person "* (Article 4 of the General Data Protective Regulation). Cookie IDs and IP addresses of mobile devices also contain location information and may point to the location of a particular person on city scale.

Revealed location information may enable identification of user's interest and preferences in different context (Sila-Nowicka & Thakuriah, 2015). Location data is considered as sensitive personal data when the context references to the physical, physiological, genetic, economic or social identity of a natural person according to the European Data Protection Directive 95/46/EC. When the location data of an individual can be linked to a certain context this may impact the privacy of an individual (van Loenen et al., 2008). For example, Jin, Long and Joshi (2012) discovered that the geographical coordinates of residential values of Foursquare users are publicly available. With the help of Google Geocoding API, it is possible to identify the full address of the venue within a range of 800 meters. The identification of the residential values of Foursquare users may lead to higher risks for privacy threats and needs to adhere to the rules in the European data protection regulation.

Another problem is that social media mine location data and sell it to third parties and do not inform their users sufficiently about the re-use of personal data. In May 2017, the Dutch Data Protection Authority (DPA) concluded that the Facebook Group violates Dutch data protection law after its investigation into processing of personal data of 9,6 million users in the Netherlands. Facebook gives insufficient information about the use of personal data. For example, the Facebook group omits to inform users that Facebook processes location data of their "friends" for advertising purposes. Besides,

the social media company uses sensitive personal data from users without their consent (Autoriteit Persoonsgegevens.nl, 2017).

Social media users should be aware that their location data is processed and re-used for diverse purposes of organizations according to Article 9 of E-Privacy Directive 2002/58/EC: *"The service provider must inform the users, prior to obtaining their consent, of the type of location data, the purposes and duration of the processing and whether the data will be transmitted to third party for the purpose of providing the value added services."* Privacy and Electronic Communications Directive 2002/58/EC, also known as the E-Privacy Directive, is the European directive on data protection and privacy rights applied to electronic communications technology and content. The directive regulates how companies should track users, collect data stored in user's devices and engage in re-use of these collected data. Location data relating to users can be only processed when they are made anonymous, or with the consent of the users to the extent and for the duration necessary for the provision of a value-added service (Article 9 of E-Privacy Directive 2002/58/EC).

In short, location information is personal data as it refers to the location of a user's mobile device, according to E-Privacy 2002/58/EC and GDPR. Apart from the location, the information also provides a context to one's behaviour and may refer to a natural person's identity in a physical, physiological or social way. The link between the location data and its context is important in order to determine whether the location data is sensitive or not (van Loenen et al., 2008). Social media informs their users insufficiently about the use of location information derived from mobile devices and profiles. This does not comply with E-Privacy Directive 2002/58/EC nor with the GDPR, because social media are obligated to ask their user's permission to analyse and process their location data.

General Data Protection Regulation

The current European directive 95/46/EC will be replaced by *General Data Protection Regulation* (GDPR), which will be fully enacted in May 2018 (de Hert & Papakonstantinou, 2016). The national regulations of EU member states regarding data protection will be replaced by the GDPR. In addition, national supervisors will be nominated for the supervision of the implementation of GDPR. In the Netherlands, the GDPR will replace the Dutch Data Protection Act (Wet Bescherming Persoonsgegevens) and the Dutch Data Protection Authority (Autoriteit Persoonsgegevens) will become a local supervisor who supervises processing of personal data to ensure compliance with the provisions of GDPR and advises on regulations in the Netherlands.

Chapter 5 Geotagging functionality in social media

The fifth chapter aims to answer second sub question 2a:

2. How do geotag features work on social media?

First, the definition of geotagging will be described, and then the geotag features on Facebook, Instagram and Twitter (Section 5.1). A technical information background will be also provided (Section 5.2).

5.1 Geotagging

Geotagging is the process of adding geographical information to digital content such as photographs, videos, blogs, and status updates on social media by web users who produce the content (Goodchild, 2007). By doing so, social media users produce location information on Web 2.0. Location information can take several shapes and forms on Web such as place names, street addresses or geographical coordinates, and can be described by different levels of spatial granularity from the country level to a postal address. Nowadays, it is also possible to check-in at companies, restaurants or any other public spaces on social media (Figure 5-1). A variety of social media are embedded with geotag features for users. One can either directly choose a place name from a list of suggestions near the located place of the device, or the user can indicate his location directly by typing it and adding it to his content. The location update of the user will be visible to the audience on the web, based on the privacy settings of the user (Tang et al., 2010).

In the Netherlands, millions of people use social media daily on their mobile devices. WhatsApp has around 10,9 million users and a relatively high number of daily users (around 7,8 million) (van der Veer et al., 2017). Facebook comes in second place with 10,4 million (active) users and 7,5 million daily users (72%), whereas Instagram has 3,2 million users in the Netherlands of whom 1,5 million (47%) post photos and view other posts daily. 871.000 users use Twitter daily which is "only" 3,4% of the total Dutch Twitter users, compared to relatively high numbers of Dutch Instagram and Facebook users. Twitter isn't as popular among users and is passed by other social media services such as Instagram, YouTube and LinkedIn (van der Veer et al., 2017).

Since 2013 the functionality of Twitter and Instagram are embedded on Facebook. Embedded posts are a way to put public posts into the content of someone's profile on social media. Users can post their Tweets and photos at the same time on Facebook as well. This means users can share their location information with the same post on several social media services at the same time. For these reasons, this paragraph focuses on the geotag functionality of Twitter, Instagram and Facebook.

Figure 5-1 Check-in at sushi restaurant in Utrecht on Facebook



Source Facebook (2016)

5.1.1 Twitter

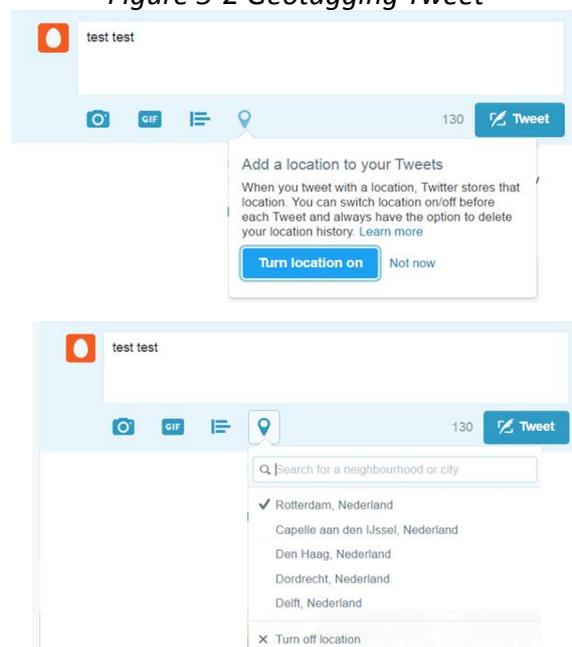
In 2006 Twitter was founded by Twitter Inc. in San Francisco, becoming publicly available the same year in July. Microblogging is the core of this social network, which enables users to write and publish short messages, also known as Tweets. The structure of social interaction of Twitter is relatively simple: a user can follow someone else on Twitter but in turn can also be followed by someone else. Tweets, which are short messages of max. 140 characters, can be retweeted by other users. Another option is to favourite Tweets of other users and to reply directly to a specific user in a Tweet. In addition, tweets may contain entities and hashtags. These so-called entities are for notifying a Twitter user who is mentioned in a Tweet. Hashtags help to associate Tweets with subjects or events. Users can also restrict the visibility of their profile and Tweets to other Twitter users. This is up to the user, as some accounts are publicly visible to everyone (Twitter, 2016). Despite the effort of users to protect their content from unwanted public, Twitter still has access to this content.

A relatively small number of tweets are geotagged (0,85%). However, this accounts for over 4 million tweets every 24 hours using an estimate of 500 million tweets per day (Twitter Help Centre, nd). There are two explicit geotagging functionalities on Twitter. When a user has opted-in, which means the requested activation of the geotagging feature is answered, one can geotag Tweets with a precise location or a place name (Twitter Help Centre, n.d). The users share their location continuously when the geotag feature is opted-in (Figure 5-2). However, it is also possible not to geotag a Tweet, even though the feature is opted-in. The first way happens with device location positioning with geographical coordinates, while the latter is a list of suggestions. This means that the user cannot specify a not yet catalogued place name (Twitter Help Centre, n.d). The places can be specified using different levels of granularity, which are denoted as country, city, neighbourhood or point of interest. Only original tweets can be geotagged. Retweets are never geotagged, because Twitter does not classify them as original content. Geotagged data contains the most information in a useful and accurate format since it contains geographical coordinates whereas profile-based locations only tells where people were born, lived as well as employed. This also depends on the geotagging behaviour of the user on social media.

5.1.2 Facebook

Facebook was launched in 2004 by Mark Zuckerberg and his fellow roommates. In the present, the company is based in California, United States. Users must register to use the social platform in order to broadcast their social life online, by creating a user profile indicating their name, occupation, work, and education. It is also possible to add fellow users as 'friends' on Facebook. Users can exchange messages, digital photos, videos and post status updates on the platform. In addition, one can also use software applications such as Candy Crush. Online users receive notifications when fellow users update their profiles or tag them in a post, acting as a calculated tool to stimulate social interaction. In addition to these posts containing videos and pictures, users can geotag their content on Facebook.

Figure 5-2 Geotagging Tweet



Source Twitter 2016

Figure 5-3 Check-in on Facebook



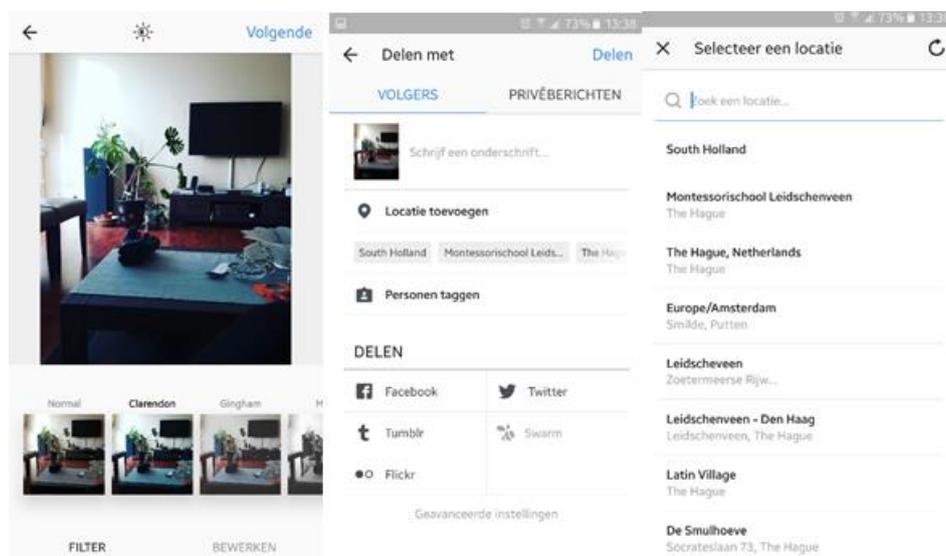
Source Facebook (2016)

There are two ways of geotagging on Facebook, comparable with Twitter. One can manually add his or her location to their text, picture or video (Figure 5-3) or choose from the list of suggestions. For example, when someone types 'Den Haag' in the application, the website returns suggestions such as "Den Haag Stad" and attractions such as "Den Haag Beach" or shops such as "Den Haag Topsport" (Figure 5-3-A). In addition, one can also choose from a list of places where he or she was checked-in in the past (Figure 5-3-B). One can also check-in at street level with suggestions (Figure 5-3-C).

5.1.3 Instagram

Instagram was founded and launched in 2010 by Kevin Systrom and Mike Krieger as a free mobile application. The application gained popularity with over 100 million active users in 2012 and over 300 million users in 2014. In 2012, the application was acquired by Facebook. On this platform, people can share photos and videos and publish the content on other social networking platforms as well such as Facebook, Twitter, Tumblr and Flickr. Instagram also offers users different manipulation tools to transform the appearance of a photo or video (Figure 5-4). Like the social interaction of Twitter, Instagram also allows the user to follow other users, called "followings" (Hu, Manikonda, & Kambhampati, 2014). Fellow users can also see the number of followers on someone else's Instagram account. In addition, users can set their privacy preferences. For example, the photos and videos of person A are only visible to his followers, and other users should ask permission to follow person A on Instagram.

Figure 5-4 Geotagging photo on Instagram with list of suggestions (Point of Interest)



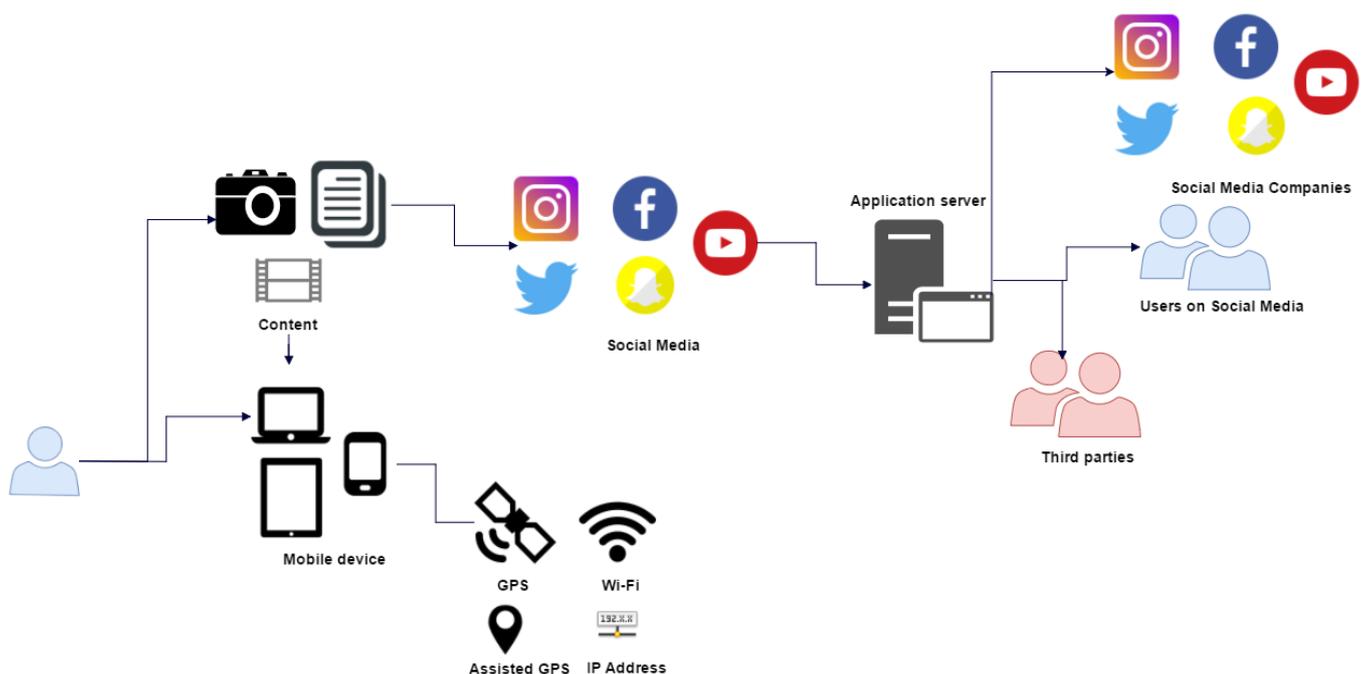
Source Instagram (2016)

Instagram users can geotag their content on the platform in the same way as on Twitter and Facebook. Hu et al. (2014) revealed that users on Instagram are much more likely to share their location compared to Twitter users. Out of 5,659,795 pictures, more than 18.8% contain location information. In addition, at least 28.8% of users have at least one of their pictures GPS tagged (97,871 out of 369,828) (Hu et al., 2014). A user can choose to add location information by checking the “add location” option, which is unchecked by default (Figure 5-6). Then, the user proceeds to use either point of interests (also known as list of suggestions) provided by Instagram or to use the exact GPS coordinates of the current location.

5.2 Technology behind geotagging

The social media user creates geotagged content with multimedia such as pictures and videos. The created content is then uploaded on social media and saved into an application server (Figure 5-5). The geotagged content contains location information positioned by the mobile device. The users on social media, the company itself and third parties have access to this information on different levels (Christin, Reinhardt, Kanhere, & Hollick, 2011). For embedding geotag functionality in an application or social network, the process requires the involvement of cellular identification or corresponding applications with GPS and Wi-Fi hotspots, assisted GPS, and Internet Protocol Address (IP).

Figure 5-5 Connection between user on social media, mobile device and identification of location



Based on Christin et al. (2011)

Cellular identification is based on the process of triangulation. The mobile device is linked to a specific base station with a unique ID when it is switched on. The base station is registered to a specific location based on the estimation of the direction from which the base station receives the signal from the mobile device (Christin et al., 2011). Although this method provides a quick positioning of a user's location, it is only accurate to approximately 50 metres in dense populated urban areas. Most social networks work with corresponding applications which were developed for a mobile operating system and only have a limited set of functionalities. These apps make use of various location APIs that are exposed by the mobile device. This could be a GPS module over mobile phone or a WiFi access point ID. The accuracy of the geographical information can range from a few meters up to many kilometres.

The quality of the geographical information depends on several factors such as the type of sensor, reference of databases and quality of reverse-geocoding of coordinates into places. GPS gives accurate positional information on four dimensions of latitude, longitude, altitude and time. The accuracy is between 4 and 15 metres (Roxin, Gaber, Wack, Nait, & Moh, 2007). Wi-Fi access point IDs can locate devices in areas that have become blanketed with public and personal Wi-Fi access points. Mobile devices can detect the unique ID from the Wi-Fi access point, and send this to a service for location identification. The MAC (Medium Access Control) address is recorded in the hardware of the device and has a unique ID for each Wi-Fi access point. The Wi-Fi access point does not provide accurate location information as GPS, but is widely used and functions well indoors. Furthermore, assisted GPS (A-GPS) is a combination of GPS and cellular identification technology. It is also a hybrid alternative solution to speed up the location identification process (Roxin et al., 2007). The information about the mobile device is transmitted through the network of base stations, which only takes a few seconds. However, most computers do not have a GPS module or cellular antenna, so the location is often identified from the current Internet Protocol (IP) address. The identification of location through IP addresses can be done with different lookup techniques such as geolocation service. This service identifies the location of a computer at different scales. The accuracy on city scale varies between 50% and 80%, while on nation scale the accuracy varies between 95% and 99%. The location identification on ZIP code level is less accurate ("How accurate is IP Geolocation?", 2017)

Chapter 6 Monetization of privacy and its risks

This chapter aims to answer the second sub questions 2b and 2c:

2b) *For what reasons do social companies collect personal data?*

2c) *What are the risks of geotagging for the privacy of social media users?*

6.1 Monetization of personal data

Personal data has become increasingly important over time for business and governmental organizations. Social media users create their content on the internet and integrate it online into like-minded groups. Users tag their friends on social media and reply to their content as well. They also reveal information about their friend's characteristics and behaviours (Tubaro, Casilli, & Sarabi, 2014). Personal data is advantageous for commercial purposes. Users reveal their personal data, which can be used for more efficient methods of online advertising and customer relation management (Tubaro et al., 2014). The storage and processing of personal and location data of social media users for commercial purposes can also be referred to as monetization of privacy (Tubaro et al., 2014).

Online advertising has been one of the fastest-growing businesses in the 21st century. By using detailed personal data, online based advertising provides efficient methods of matching advertisers and consumers. According to Evans (2009), the matching can be achieved in two ways: content creation that facilitates the aggregation and sorting of potential customers, and behavioural targeting (as cited by Tubaro et al., 2014). Behavioural targeting refers to sorting and identifying potential buyers based on observation of individuals' characteristics and behaviours such as age, gender and location (Tubaro et al., 2014).

More and more companies make use of web-based communities and peer-to-peer collaboration tools as an extension of traditional customer relationship management. For example, the Facebook pages of companies such as [Hunkemöller](https://www.facebook.com/hunkemoller/)² or [Albert Heijn](https://www.facebook.com/albertheijn/)³ engage in this behaviour. Companies and non-commercial organizations are aiming to connect with their customers or promote new products and services online. Companies can rely on existing services without designing their own and users do not feel the pressure to maintain more accounts and profiles (Tubaro et al., 2014). Social media also offers new opportunities for marketers and other companies by enabling word-of-mouth mechanisms to be exploited by advertisers. Facebook has devised several ways to target consumers based on the choices and behaviours of their friends. Several external services have been using Facebook identifiers for logging in or the infamous "like" button for external websites. There are also risks bounded to these online customer relationship management practices. When companies use social media platforms or connect their private network to social media, any personal information has greater potential to leak to a wider set of connections.

Social media companies have an intermediary role between the social media users and (advertising) companies who benefit from the personal data. Such companies must deal with complex challenges involved in the market. Most of the popular social media services have been free of charge for users, while advertisers must pay for commercials on these services. The "free" services of SM to its users attracts huge numbers of people, thereby increasing the value of advertising space and leading to higher prices for advertisers. With the value extracted from advertising fees, social media companies can improve their services and attract even more users for monetization of personal data (Tubaro et al., 2014).

² Hunkemöller Facebook <https://www.facebook.com/hunkemoller/>

³ Albert Heijn Facebook <https://www.facebook.com/albertheijn/>

6.2 Collection of personal data and location data

Any piece of content users voluntarily or unintentional disclose on Instagram, Twitter, and Facebook, becomes publicly available as it is controlled by the user's privacy settings. The privacy policies of all three social media services state that the user has their own responsibility to make decisions regarding whether to disclose personal information and share content on their social networking platform. The privacy policy of Twitter mentions that the users alone are responsible for the posting of their own Tweets and other content they submit through the services. The users have their own responsibility to protect their own privacy. The reader of the privacy policy is frequently warned for the fact that Tweets are publicly visible to others: *"Twitter broadly and instantly disseminates your public information to a wide range of users, customers, and services, including search engines, developers, and publishers that integrate Twitter content into their services, and organizations such as universities, public health agencies, and market research firms that analyse the information for trends and insights. When you share information or content like photos, videos, and links via the Services, you should think carefully about what you are making public."* (Twitter Privacy Policy, n.d.)

Besides the warnings, the general argument behind the urge for collecting all different kinds of information in big amounts is repeated through the policies of all three-social media services. The general argument is as follows: To improve and personalize the services with more relevant content like local trends, stories, advertisements and recommendation mechanisms such as suggestions to follow other users on Instagram and Twitter, or to invite friendship request on Facebook.

Collected data and the purposes

Social media companies gather, sort and repack the information of users in a way that is relevant for advertisers (Tubaro et al., 2014). The privacy policies state what personal data the social media company gathers and disseminates to other parties. Social media collects personally identifiable information about users such as names, contact information, and locations. This information may be categorized into several sub groups. First, twitter collects information from users upon sign-up such as name, username, contact information and address books on email accounts.

Second, metadata provided with Tweets or photos on Instagram are also collected and saved by the server. Metadata contains technical data about how, when and by whom a piece of content was collected and how that content is formatted (Instagram Privacy Policy, n.d.). When the metadata also contains a hashtag, geotag, comment or other kind of data it makes created content more searchable by others. When the photo or video is geotagged, the latitude and longitude will be stored with the content and be searchable on for example Instagram and related API's.

Third, the metadata and the posted content itself may contain location information. The user may disclose the location information manually or with points of interest alongside the Tweet. The positioning of the location may also be determined with other techniques such as GPS on mobile devices, Wi-Fi hotspots, cellular identification or IP address (Twitter Privacy Policy, n.d.).

Fourth, social media like Twitter or Instagram also collects information about one's browsing activity via cookies and similar technologies (Twitter Privacy Policy, n.d.). The purpose of cookies is to better understand how the user interacts with the services of Twitter, to monitor aggregate usage by the users and the routing web traffic. However, Twitter does honour users' decisions to use a Do Not Track browser option. Instagram also make use of cookies and similar technologies like pixels, web beacons and local storage to collect information about how an individual uses Instagram. Advertisers and other partners are also allowed to serve advertisements or services to users, which is based on cookies.

Fifth, log data is collected when one visits a social media platform. This data contains information about

a user's IP address, browser type, operating system, browsing history and location. Twitter and Facebook uses log data to make inferences for customized content and advertisements (Twitter Privacy Policy, n.d.). Instagram uses third-party analytic tools to measure web traffic and usage trends based on the log data (Instagram Privacy Policy, n.d.). Instagram claims the information from the analysis assists with improving the service. Facebook also collects information about how the user browses and uses their services along with attributes of the operating system, location information and connection information (Instagram Privacy Policy n.d.; Facebook Data Policy, n.d.). This information provides reports or personalized content and advertisements. Additional data files are the device identifiers which are stored in or associated with mobile devices. A device identifier may be stored in connection to hardware or with the device's operating system.

Privacy policies do not provide a complete overview of which personal data is mined from social media platforms. It is also difficult for users to get an overview or receive feedback regarding their own personal data that is collected by social media companies and advertisers (Martijn & Tokmetzis, 2016). In 2011, the association Europe versus Facebook, founded by Austrian Max Schrems, filed many cases with the Irish Data Protection Commissioner (DPC), because Facebook failed to comply with the rule of providing feedback to its users with their own personal data when requested to do so (Europe versus Facebook, 2017). Because this case Facebook released a part of its data pole that are held by the company which contains more than 57 personal data categories. At least 9 of the 57 categories contain location information (Table 6-1). Facebook gathers location information of their users and their devices through several ways. For example, through IP-addresses and geotagged content such as photos and updates on profiles.

Table 6-1 Types of location data collected by Facebook

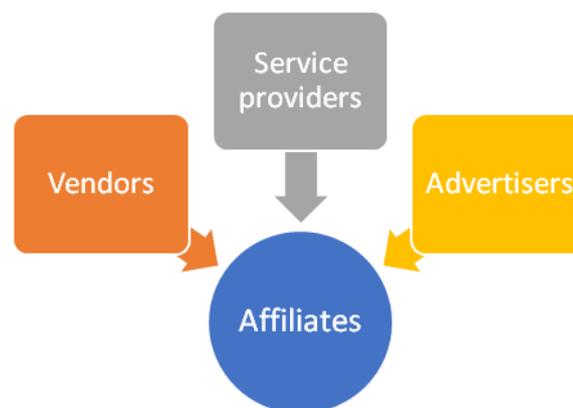
Category	Description
Address	The address typed by the user. Though it is unclear if and where Facebook gathers information about the user from other users or through their mobile device.
Check-in	Lists of all checked-in places in the past. This data set also consists of the author, other tagged users, personal messages and an exact latitude, longitude and altitude. Facebook also adds an individual ID number and an exact time stamp alongside every check-in.
Current city	The city with the ID number where the user currently lives
Events	The invitations on Facebook for events also contain location information of the place where these events will happen. The lists for this category contain all events the user has ever has been invited to, disregarding rejection or non-reaction from the user's side.
Hometown	Home town with an ID number
Last location	Although it is unclear how Facebook precisely gets this information about the user, it contains the location of the user. According to "Europe versus Facebook" it might be a mixture of check-ins, location information from applications, or other kinds of geotagged content on Facebook. Last used IP address.
Networks	This data set contains the networks the users are member of. Those networks might be specified with location information.
Photos	Whether the photo is geotagged on Facebook, one's uploads on Facebook contain location information. This location information comes along with the users' mobile device which has been used for taking the picture.
Real time activities	The tracking Facebook does on its own page. It also contains IP address.
Recent activities	These data sets show all the log-ins on Facebook and contain all IP addresses and cookie information as well as information about locations and time.

Third party use

The definition or the conceptualization of the ‘third party’ is lacking in privacy policies. The privacy policies of Facebook, Instagram and Twitter do not mention which companies are considered as partners. Instagram claims they do not rent or sell personal information to third parties outside Instagram without the consent of the user (Instagram Privacy Policy, n.d.). Though there are some remarks regarding third parties. The personal information of users may be shared with businesses that are legally part of the same group of companies that Instagram is part of. These kind of partnerships are called “Affiliates” (Instagram Privacy Policy, n.d.) Another group whom Instagram provides information to are the Service Providers. Service Providers are third-party organizations that provide Instagram services such as analysis and technical support. Advertising partners also receive personal information of users. Twitter can sell or share users’ information with third parties. Though there are some restrictions applicable for the advertisers concerning sensitive subjects: “ *Our Twitter Ads Policy also prohibits advertisers from targeting ads based on categories we consider sensitive, such as race, religion, politics, sex life, or health. If you prefer, you can uncheck the Promoted Content setting within your Security and Privacy Settings so that your account will not be matched to information collected by ad partners, or by us directly on those partners’ websites or apps, to tailor ads to you.*” (Twitter Privacy Policy, n.d.)

Facebook also cooperates with third parties such as advertising and analytics services. Facebook claims they do not share identifiable information such as name or email address unless the user gives Facebook permission. However, problems occur when the user doesn’t give permission to collect information from the user’s profile. The applications from third party developers will not work or the user will not be able to profit from the provided services. Just like Instagram and Twitter, Facebook also transfers users’ personal data to vendors, service providers, and other partners who support Facebook’s business for analysing, technical support and measurement of effectiveness of ads (Facebook Data Policy, n.d.)

Figure 6-1 Types of affiliates



6.3 Dimensions of location privacy problems on social media

Geotagging functionality operates in social media and creates location data. Social media is dependent on users’ mobile devices for acquiring the current location by using GPS, WiFi triangulation, or cellular networks. Along with the combination of other personal data it may significantly increase risks of interference with the right to privacy. If there is a minimal privacy securing mechanism it is possible to sketch a highly-detailed user profile, track and predict the user’s daily movement, and their behaviour. Companies, governmental organisations, and hackers may misuse this data for economic

gain, physical stalking, or to gather unjustified legal evidence (Puttaswamy et al., 2014). The attacker can easily obtain location information based on check-ins and geotags from web pages, extract users' points of interest from collected data, or even automatically transform the place name to a GPS coordinate or vice versa (Li, Zhu, Du, Liang, & Shen, 2016). Alrayes and Abdelmoty (2014) identified four dimensions of problems that may affect location privacy of mobile device users: (1) amount of collected data and its quality, (2) accessibility to location information, (3) exploitation of location data and (4) security of location data.

Amount of collected data and its quality are divided into three aspects: method of collection, types of data and data volume. Method of collection refers to the mode of data collection. The data collection may happen in different modes and in different time blocks (Alrayes & Abdelmoty, 2014). When the mode of data collection is automatic it can be continuous due the default option in settings. It may also be manual at periodic times when a particular user checks in occasionally on Facebook. The mode of data collection will impact the volume of collected data and its accuracy (Alrayes & Abdelmoty, 2014). The volume of collected location data is dependent on user attitude and behaviour when using the application (Alrayes & Abdelmoty, 2014). The pattern of data and the frequency of usage will determine the density of the data over time. This may influence the type of information that may be inferred from the collected data. Mobility patterns, social relationships and such can be studied from the data.

There are **three types of data** that can be associated with location data: spatial semantics, non-spatial semantics, temporal semantics. *Spatial semantics* refers to types of information that can be used for identification of places (Alrayes & Abdelmoty, 2014). The data may have latitude and longitude coordinates or contain place names or street addresses. Instagram allows users to geotag their photos using the Foursquare API ("API Endpoints • Instagram Developer Documentation," n.d.). Twitter uses Google API to select place names with a location on a map. The detailed and accurate places are linked with the users. *Non-spatial semantics* are types of data about the user and places that are associated with location information such as reviews, tags and pictures. This type of data may contain personal information about the users. *Temporal semantics* represents the time of a visit and the duration of the users visit (Alrayes & Abdelmoty, 2014). In social media, the time of visit is registered by users when they check-in at a place. The actual GPS coordinates of the user's device may validate the user's physical presence in the place.

Accessibility to location information refers to how much of the user's data is available and visible to the user, other users and third parties of the user (Alrayes & Abdelmoty, 2014). In general, users of location-based services and social media have limited access to their collected data which contains location information and other kind of personal data such as name and friends list. The Application Programming Interfaces (API) provides access to all publicly available user information to third parties. The content of the users is publicly available by default unless the profile on social media is private.

Location data exploitation refers to how the application or third parties can utilize the data and for which purposes which involves the exploitation of users' location and other personal data (Alrayes & Abdelmoty, 2014). This may lead to various levels of privacy threats. People have regular routines and be characterized by a set of significant places in their daily routine, which makes it possible to identify a user from his/her mobility data. There are series of techniques for identifying individuals from their GPS movements. Rossi, Walker, & Musolesi (2015) provide a detailed analysis of the discriminatory power of speed, direction and distance of travel. With simple, yet effective techniques one can identify users from location information (Rossi et al., 2015), although there might be some differences between identification based on GPS movement and geotagged content on social networks. Check-ins at users' residential venues have more risks to be identified by hackers, whereas check-ins to a popular restaurant or public transport are less discriminative, since these places are likely to be visited by a

crowded public with different users with similar check-in patterns. Rossi et al. (2015) investigated the interdependence between location semantics from check-ins and privacy by studying the relationship between the characteristics of a venue and the ability of an attacker to discriminate between the identities of different visitors of that venue. The findings of Rossi et al. (2015) shows that frequency of visiting a venue or frequency checking in does not determine the difficulty of identifying a person. Rather, for the identification of a user the type of venues and check-ins matters.

Designers or developers of location-based social networks should consider the discriminatory power of categories when the privacy policy and **security** is implemented (Rossie et al., 2015; Alrayes & Abdelmoty, 2014). The security is related to the level of data protection provided by the application for securing data against risks of loss or unauthorized access (Alrayes & Abdelmoty, 2014). Data protection can be provided by a concept called *privacy by design*. Sensitive personal information is considered as early as possible in the design phase and security technologies are implemented to protect personal data.

In short, social media companies collect user's personal data for monetization. The monetization process works in three ways. Firstly, social media provide data to their affiliates for improvement of their services on their platforms such as personalized advertisements and recommendation mechanisms. Secondly, the personal data is used for online behavioural targeting by advertisers. Advertisements influence the behaviour of social media users. Thirdly, companies and non-profit organizations manage their customer relationship through online communities on social media. The free services of social media attract a large number numbers of users, whereas advertisers and companies have to pay for advertising on community platforms.

The more users a social medium has, the more the value of advertisements leads to higher prices for marketers. A larger number of daily visitors on your social medium will lead to a higher ad revenue, as a large audience increases the price of advertising for marketers. Social media looks for ways to improve their services and monetization of personal data for advertisement purposes.

The risks of geotagging or sharing location information are grouped into 4 dimensions that could affect users' location privacy. First, the amount of collected data and its quality defines the level of detailed profiling. Second, the type of location data (spatial semantic, non-spatial semantic and temporal semantic) provides the accuracy of the location information. Third, the accessibility of? data with location information refers to how much of the user's data is available and visible to others. Social media users have limited access to their own collected data. However, APIs provide full access to publicly available data of users. This accessibility also creates the possibility for exploitation of location information by third parties. Fourth, lack of security and data protection may also exacerbate misuse of personal data. Therefore, developers of location-based social platforms should consider the risks of privacy threats and provide security to protect data and improve privacy policy related to re-use of location information by third parties.

Chapter 7 Privacy paradox: behaviour, attitude and concerns

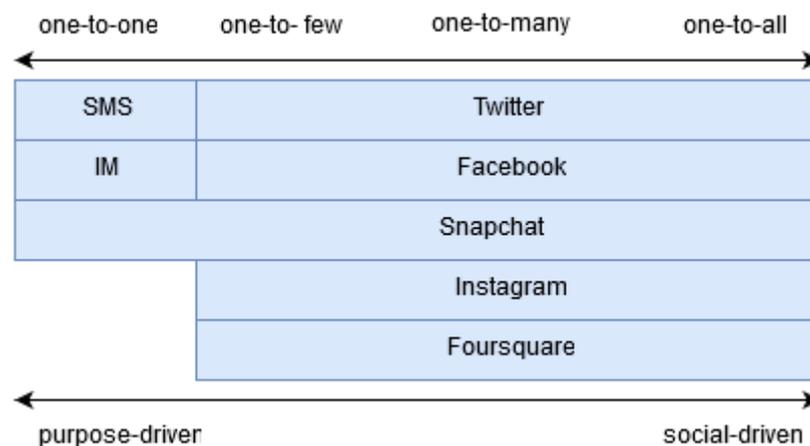
The seventh chapter is a theoretical chapter that lays down the basis for the online survey, and aims for the third and fourth sub questions:

- 3) *What is the self-reported geotagging behaviour on social media?*
- 4) *What are the attitudes and concerns of users on location privacy when they are using their social media profiles?*

7.1 Behaviour

Location-based technologies create opportunities to develop interactive experiences that rely upon human movement behaviour and experiences of places (Dourish, 2007). Users can discover new places and share their experiences of places with other people. There are several reasons why an individual is willing to share her or his whereabouts with fellow users on social media. Disclosure of location information might be a social, emotional and moral way to express moods, lifestyle and events (Barkhuus, 2008; Cramer et al., 2011). Users consider between preservation of privacy and the benefits they expect to gain by sharing their location information with fellow users on web. They also search how to balance private and public spheres (Olteanu, Huguenin, Humbert & Hubaux, 2016). The decision whether to disclosure location information also depends on the undertaken activity, its location, and the potential audience on social media.

Figure 7-1 Motivation for location sharing and recipient size



Based on Tang et al. (2010)

The geotag functionality gives the users an opportunity to self-report their location, and to decide who can view their whereabouts (Lindqvist, Cranshaw, Wiese & Zimmerman, 2011). According to Tang et al. (2010) location sharing in social networks has transformed from purpose-driven sharing, which is done in response to specific location request, to social-driven sharing, which is done to large social groups (Figure 7-1). The former happens for mainly pragmatic reasons, while the latter is more for promoting and sustaining social capital within a network. Most of the time, purpose-driven sharing applies to one-to-one communication, while social-driven sharing makes most sense in the context of social media with one-to-many communication. The disclosure of location information on social media can be traced back to the “social-driven location sharing” theory (Lindqvist et al., 2011). Social-driven sharing is often used to boost self-representation and to attract attention from fellow users on social media (Tang et al., 2010). However, one-to-many communication may not be as simple as it seems. For example, on Twitter, the communication involves more complex reasoning than one-to-one or one-to-

few communication between users whereas the location information is only visible to a small audience. Benisch et al. (2011) found that participants were comfortable sharing location 93% of the time to friends and family and 60% of the time with friends on Facebook. Most users are willing to share their location information to peers such as co-workers, friends and family, but they explicitly refuse to share their location information with strangers.

Participating in online social networking may help to increase social capital, to increase a sense of connectedness with fellow users and to work on self-representation on web (Table 7-1). Cramer et al. (2011) explored the motivations for geotagging by the users of Foursquare, Facebook and Twitter. Users check-in at Foursquare to receive discounts on shops, to discover new venues and to meet new people. On Facebook and Twitter, users were more concerned about their online behaviour rather than being concerned with their online privacy. The expected benefits are related to relation development and self-development such as social exchange, establishing one's image in a community and maybe even the ability to influence fellow users on social media on their attitude or thoughts on a subject (Lee et al., 2013; Pötzsch, 2009) (Table 7-1).

Table 7-1: Expected benefits from disclosure personal information

Expected Benefit	Description
Relation development	Social exchange, serendipity, collaborations
Self-presentation	Establish image, reputation
Self-clarification	Understanding oneself, thinking about own situation
Social validation	Attaining approval, being justified
Social control	Influencing others' attitude/behaviour/opinion and thoughts

Source Pötzsch (2009) and Lee et al. (2013)

Type of places

The type of places also matters to the users in their decision whether to disclose their location information on web. The selection of location is based on its popularity and if it's private or public. The more popular a place is, the more likely other people will go to visit it (Hasan, Zhan & Ukkusuri, 2013). People prefer to shop and dine at places, go to restaurants after shopping and check-in to universities mainly on the weekdays (Long, Jin & Joshi, 2012). Location sharing patterns reflect the daily life of people (Dourish, 2006). Most of the time, users are more eager to share their vacation places, restaurants, bars and daily trips, whereas the location of their residential addresses is more likely perceived as sensitive information (Wagner et al., 2010). Wagner et al. (2010) suggest there is a strong hierarchical distinction in how users choose to disclose their location when they are at home. Users are less willing to share location of more less private places, places that are not visited uniquely like their home, or their family's home. Conversely, public places, just as restaurants are considered less private, lead to users being more likely willing to share these kind of places (Toch et al., 2011).

7.2 Concerns

Users are, in general, privacy-aware to some extent and have knowledge about numerous controversies around privacy such as the continuous changes in the privacy settings on Facebook or the concerns around location tracking on smartphones (Zafeiropoulou et al., 2013). When a user decides to disclose personal information on a social network, this person also weighs the expected benefits and expected risks of his or her online behaviour (Table 7-1 and Table 7-2). Lee et al. (2013) investigated what kinds of risks and benefits exist among users when one shares personal information. The expected risks are related threats such as identity theft, surveillance and stalking (Table 7-2). In addition, users may get a negative reputation because of their behaviour on social media or even their job or position may be jeopardized. The expected benefits and risks influence users' intention to share their personal information on web. The effect of expected benefit was found to be stronger than that of expected risk (Lee et al., 2013).

Table 7-2 Expected risks from disclosure personal information

Expected Risks	Description
Face risk	Losing face on social network, negative reputation or embarrassment
Relational risk	Jeopardizing friendship or relationships,
Security risk	Identity theft, stalking, kidnapping, surveillance
Role risk	Jeopardizing job, position and role
Stigma risk	Being disgraceful, immoral and unaccepted

Source Pöttsch (2009) and Lee et al. (2013)

7.3 Attitude

Privacy interest positions

When individuals are uncertain about their preferences they often search for cues in their environment to provide guidance for decision making. Cues are both a function of a given context as well as behaviour. Individuals can exhibit what ranges from extreme concern to apathy about privacy. This attitude towards privacy depends on a given situation and its context. The definition and meaning of privacy differs between persons (Acquisti, Brandimarte, & Loewenstein, 2015; Westin (2003) as cited by van Loenen et al., 2008).

In other words, the sense of what must be kept private differs from person to person. Harris & Associates & Westin (1995) (as cited by van Loenen et al., 2008) mentions three ideological-interest positions on privacy: *privacy fundamentalists*, *privacy pragmatists* and *privacy unconcerned*. People who are *privacy fundamentalists* perceive privacy as a right of autonomy, whereas *privacy pragmatist* value privacy as a right to seclusion. Privacy is valued, but certain trade-offs are acceptable regarding the expected benefits from sharing personal information with applications and others. Individuals, who are *unconcerned* about their privacy, perceive it as a property right. They assign a lower value to privacy claims than business efficacy. One may even perceive societal-protection interest and governmental intervention as unnecessary and costly. Most people are privacy pragmatists since they are willing to

trade their personal information for other benefits such as discount, self-representation or discovering new venues (Acquisti et al., 2015; Lee et al., 2013).

Context-dependency

People may also seek privacy in public by managing their privacy (Acquisti et al., 2015). For example, users on social media share personal information with their followers and friends, while most users deny access from their profile to strangers. Privacy preferences and attitudes are not static, but dynamic as it depends on the external factors that become apparent in each context (Zafeiropoulou, Millard, Webber, & O'Hara, 2013). Nissenbaum (2011) concludes that social expectations of humans affect their beliefs regarding what is private and what is public. Such expectations vary with specific contexts. For example, visiting a hospital may be considered as more sensitive than visiting a restaurant because the former deals with health concerns while the latter is a social activity (Li et al., 2016). Users consider the expected benefit from sharing personal information, which can be used for services such as discounts or benefits from services. This decision-making regarding privacy and gaining benefit from it is also known as privacy trade-off (Lee et al., 2013).

Default settings and interface design

Our perception and attitude towards privacy seems to be influenced by the development in social networking nowadays (Tubaro et al., 2014). Users negotiate their attitudes towards privacy in response to new forms of interactions and experiences offered by positioning technology embedded in social applications.

Attitude towards privacy can be also influenced by default settings, malicious interface design, antecedents, controlling and the degree to which social media policies are transparent. Social media companies and advertisers have developed an economic interest around personal data. Therefore, some entities, who are interested in personal information, have also developed expertise in exploiting behavioural and psychological processes to promote disclosure of personal information. These efforts depend on the malleability of privacy attitudes and concerns of social media users (Conti, Point, & York, 2010). This malleability refers to the fact that some factors can be used to activate or suppress privacy concerns, which in turn influences behaviour (Acquisti et al., 2015).

Many social media platforms encourage their users to share their context with others as part of their interface (Piwek & Joinson, 2016). For example, Facebook asks their users question such as "What are you doing right now?" while Twitter asks "What happened?". Context refers to the surrounding situation of a user and contains features such as the location, emotion and presence of others. A study by Tang, Lin, and Hong (2010) has shown social media users feel more comfortable with sharing location information when there is a choice of different privacy settings and different location granularities within geotag functionality. This means users can choose between location information based on diverse scales such as region, city or even a venue. People are more likely to be eager to share their location information when the social media platform gives more location granularity options (Tang et al., 2012). Users prefer to use semantic names to regulate privacy by not making them directly locatable. Users do not prefer to share exact location. The urge gets stronger to manipulate the granularity of location sharing, when a user disclosure its whereabouts to less familiar users (Lin et al., 2010). Privacy configurations that support varying location granularities may change how privacy rules are defined and under which circumstances locations are shared by users. More abstract location descriptions can lead to more open location sharing along less complex rules and fewer negatively phrased rules according to Tang et al. (2012).

Default settings are used by different entities to affect information disclosure from users. Sticking to default settings is convenient, because people perceive default settings as implicit recommendations

(Acquisti et al., 2015). Hereby, the default settings affect the visibility of one's profile on social media, or the possibility to opt-in or opt-out on a website's privacy settings (Acquisti & Grossklags, 2005). Malicious interface designs can be used as well in order to confuse users and make them disclose more information (Conti et al., 2010). Meanwhile, antecedents affect concerns and can be used to influence behaviour such as users' trust in the entity. Receiving one's personal data soothes concerns. In a study by Hoofnagle & Urban (2014), 62% of respondents to a survey believed that the privacy policy implied that a site could not share their personal information without their permission. This suggests that users do not read policies or misinterpret the policies (Hoofnagle & Urban, 2014 as cited by Acquisti et al., 2015).

The user's control over personal information is another feature that can be misused to create more trust among users. Users may have the feeling that they are in control of their own personal information flow by managing the other user's accessibility to their profiles and content. This doesn't change the fact that social media companies still have the access to mine personal data from their users. The transparency of companies' data practices may soothe the privacy concerns. However, it can be easily rendered ineffective. As mentioned before, the majority of web users do not read privacy policies, but nearly half of the sample described online privacy policies as difficult to understand due language use (Jensen & Potts, 2004, as cited by Acquisti et al., 2015).

User characteristics

The characteristics of a user appear to have an impact on their willingness to disclose personal information and their attitude and concerns as well (Li & Chen, 2010; Taddicken, 2014). Characteristics such as gender and age have been studied by academic researchers, whereas education and its influence on privacy concerns have been less investigated (Bergström, 2015). According to Blank et al. (2014) users with low education tend to be less concerned with privacy risks, and more highly educated users are more likely to utilize privacy protection (as cited by Bergstorm, 2015).

Gender has been proven to influence the relation between privacy concerns and general willingness to share personal information, as it seems female users are in general more willing to share personal information on social media. Female users seem to be more self-regulated to protect their privacy, and are more privacy-aware (Li & Chen, 2010). Women are more likely to share their interests and other personal information, but are more careful sharing sensitive personal information like their telephone number (Tufekci, 2007). They are also more cautious granting access to their information to fellow users on social media (Fogel & Nehmad (2009), as cited by Taddicken, 2014). However, according to Taddicken (2014) the gender differences only exist regarding accessibility to sensitive personal information. The study of Jin et al. (2012) shows contradicting results from their survey. Female users of Foursquare are more likely to expose their check-ins at residential venues compared to male users on the application. As it seems, the effect of gender on privacy concerns has been somewhat inconclusive (Yao et al., 2007, as cited by Bergström, 2015). Women tend to report higher levels of concern than men do, however this seem to be a trend rather than statistically significant in researches.

The concerns of privacy increase with age, from teenage years to middle age. Li & Chen (2010) suggest older users may have more stable social relations with friends and families and therefore may prefer to share personal information with known connections rather than strangers. Older users show a more protective attitude towards privacy, whereas younger users were more likely to be better at managing privacy settings (Blank et al., 2014, as cited by Bergström, 2015). The differences in privacy concern between young and adolescent users were mediated by differences in privacy conception (Steijn, Schouten, & Vedder, 2016). Adolescent users are more likely to associate situations related to personal information. Furthermore, adolescent users have a different notion of privacy, and contrary to older

users they do not consider personal information such as age, relationship status or sexual orientation to be private, and see this as a less prominent aspect of their privacy conception. Another reason could be that younger people are more accustomed to socializing and disclosing their personal life on such social platforms. In addition, they may be more aware of the functionality and potential perils of social platforms. Younger individuals were more concerned with their location sharing behaviour, and they had less trust in social networks (Thomas et al., 2013). An explanation could be that younger users grown up in the information age with mobile devices may be more aware of their functionality and potential perils (Rahman, 2012).

The characteristics of social media users who use location services are also studied. For example, Sloan and Morgan (2015) identified the demographic characteristics of Twitter users by analysing two different datasets, collected from Twitter with Twitter API and differentiated between those who enable location services and those who do not. They investigated how gender, age, class and language are associated with the behaviour of geotagging tweets and enabling location services. There appear to be statistically significant differences for both behaviours for all demographic characteristics. There are also significant demographic variations between the users who opt in to geo services and those who geotag their tweets (Sloan & Morgan, 2015). Female tweeters are more likely to enable location services while males are more likely to geotag their tweets. The differences in age are significant, but relatively small (Sloan & Morgan, 2015).

Although Twitter users who geotag their Tweets are not representative of the wider Twitter population (Sloan & Morgan, 2015), the behavioural difference related to gender and age may be significant. However, the differences in socioeconomic status, location and education can be sizeable between the groups (Sloan & Morgan, 2015). Inequalities in education and socioeconomic status affect the degree of people's web skills, including the ability to understand privacy settings and to adjust and fine-tune them to one's preference.

Hence, disclosure of location information is a selective outcome of complex decision-making involving several factors such as the type and size of the audience and the type of place to be shared. The expected benefits such as monetary benefit, gain of social capital and privacy considerations also have an influence on the decision-making. Users do have their preferences regarding their use of social media and managing their location privacy.

Chapter 8 Sample

Before the results are presented in the next chapter, 9, the survey sample will be described in this chapter. The demographic characteristics of the participants will be described (Section 8.1) and the response rate of the questions will be analysed (Section 8.2). The representativeness of the sample will be also discussed (Section 8.3).

8.1 Characteristics of the participants

The survey was conducted online from 13 March until 7 April 2017. With the snowball method, the survey was spread online via Facebook, Instagram and WhatsApp. To be sure the survey reached every user from all ages the participants were asked to spread the survey in their own social network.

Out of 181 participants, 92 were females (50%) and 89 were men (48,4%). The average age of the participants is 27 years with a standard deviation of 10,45 years. The ages are reclassified into classes and divided by gender (Table 8-2)⁴. Most female participants (70,7%) are classified in the age class " 20 to 40 years ". It is apparent that female users are presented at a higher amount in the younger age classes than in the older age classes (Table 8-2). The highly educated participants are overrepresented whereas low educated participants are underrepresented at 1,7% in the survey sample (Table 8-3)⁵.

Table 8-1 Age of the participants

Age classes	Man		Female		Total	
	Count	%	Count	%		
Younger than 20 years	14	15,7	17	18,5	31	
20-40 years	61	68,5%	65	70,7	126	
40-65 years	14	15,7%	10	10,9	24	
65- 80 years	0	-	0	-	0	
80 years and older	0	-	0	-	0	
Total		89	100%	92	100%	181

Table 8-2 Education level of the participants

Education level	Man		Female		Total
	Count	%	Count	%	
Low	3	1,7	0	-	3
Middle	24	13,3	43	23,8	67
High	62	34,3	49	27,1	111
Total	89	49,2	92	50,8	181

8.2 Response rate

The non-response rate can't be estimated due the nature of distribution of the survey. The

⁴ The classification of age groups is based on the classification of Centraal Bureau Statistiek (CBS).

⁵ The classification of education level is based on the definition of [CBS](#).

survey was accessible via social media and there was no mailing list. Despite the lack of non-response rate, it is possible to calculate the non-response rate of the questions of the survey. Using the “Missing Values Analysis,” it is possible to generate an output for identifying patterns in missing values in several variables.

In total, 184 social media users filled in the survey online. Missing Values Analysis (MVA) shows that 105 (57%) cases are filled in, whereas 79 (43%) participants didn’t fully complete the survey. There are different reasons as to why a respondent didn’t finish the survey or answer all questions: due to refusal, limited time or losing interest.

With the help of MVA the non-response rate is calculated for each question⁶. Table 8-1 illustrates the summarized value counts and the missing values in each section⁷. The non-response answers were most common in the questions related to privacy attitudes with 26,6% missing values of N=184, changing privacy settings of social media with 29,9% missing values, and the Likert-scales about the concerns regarding location privacy with 25,7% missing values (Table 8-1). The multiple response questions about audience on social media and the re-use of location information by third parties show a steady non-response rate. All items of both questions have 139 valid cases while 45 (24,5%) values are missing.

Table 8-3 Response rate questions

Survey questions	Items	Valid	Missing		
		N	Count	Percent	
1	Gender	181	3	1,6	
2	Age	179	5	2,7	
3	Education	181	3	1,6	
4	Social media*	156	28	15,2	
5	Mobile Device *	155	29	15,6	
6	Geotag functionality	157	27	14,7	
7	Content *	157	27	14,7	
8	Location *	161	23	12,5	
9	Motivation	116	68	36,9%	
10	Change settings	132	52	28,3	
11	Satisfaction settings	129	55	29,9	
12	Audience on social media	Home*	139	45	24,5
		Workplace*	139	45	24,5
		Trip *	139	45	24,5
		Hospital	139	45	24,5
		Political event *	139	45	24,5
13	Re-use of location information	Social media *	139	45	24,5
		Advertisers *	139	45	24,5

⁶ There are two kinds of missing values: systematic missing values which are unanswered questions and discrete missing values such as “I don’t know” answers to multiple choice questions or Likert-scales. The first type of missing values is labelled as Type A and B, and are defined with the numbers “999” and “0”. The second type of missing value is type C and is labelled into two kind responses: “Weet ik niet” or “niet van toepassing”.

⁷ The questions with * are divided into several items in the survey. Each item has the same valid cases and amount of missing values. To abridge the large table, which is available in Appendix 2, the mean is calculated for each question.

	Companies *	139	45	24,5
	Research & Universities *	139	45	24,5
	Government *	139	45	24,5
	Intelligence services *	139	45	24,5
14	Attitude	135	49	26,6
15	Location privacy concerns*	136	47	25,7

8.3 Representation analysis

Using the representation analysis with the distribution of categorical variables, one can determine if there is a selective non-response in the sample. The distribution of the categorical variables gender, age and education within the sample is compared with the population of the Netherlands. The survey was aimed to all social media users in the Netherlands with diverse age classes and educational backgrounds.

The distributions of these variables in the sample are compared with the demographic data from CBS. With the help of Chi-Square goodness-of-fit test the distribution of the sample and population are compared with each other (Vocht, 2011). If the differences are not significant the sample is considered to be representative for the population. However, if the differences are significant for some groups, such as the underrepresentation of lower educated or elderly people in the sample, the classes might be weighed in the analysis on the condition that differences between the sample and population are not too big (Vocht, 2011).

Gender is one of the most common categorical variables to test the representativeness of a sample. The ratio between males and females is compared between the ratio of gender in the survey sample and in the Netherlands (Table 8-4). The zero hypothesis is that the quota male and female are the same in the survey sample, as well as in population. The asymptotic significance level is 0,908, which means the zero hypothesis is accepted because the asymptotic significance level is higher than $\alpha=0,05$ (Table 8-5). The sample is representative regarding the gender ratio in the population.

However, the age classes and education level are not representative for the population of the Netherlands (Table 8-7 and Table 8-9). The age distribution of the sample is divided into 5 classes, based on the classification of CBS. As mentioned before, the younger people are over presented whereas the elderly people are underrepresented. Social media users who are 65 and older are not represented in the sample (Table 8-6). This might be a problem for the Goodness-of-fit test since all classes should have values to meet the requirements of the test. To "solve" this problem only the classes "Younger than 20 years", "20-40 years" and "40-65" are used for the representative analysis (Table 8-7). Although this might not be correct, it will provide a non-complete insight regarding the age classes. The zero hypothesis states that the ratio between the age classes of the survey sample are like the ratio of age classes in the population. The asymptotic significance level is 0,000, which means the zero hypothesis is rejected because the asymptotic Significance level is lower than $\alpha=0,05$ (Table 8-5). The sample isn't representative regarding the age classes' ratio in the population.

Education level is a suitable categorical variable to test the representativeness of a sample. The ratio between the education levels are compared between the survey sample and the Netherlands (Table 8-8). The ratio between the education classes of the survey sample is not similar to the ratio of the population. The participants with low education level are underrepresented in the sample (1,7% compared to 32,1%), whereas the highly-educated participants are over represented (61,3% compared to 27,9%) (Table 8-8). The zero hypothesis is that the quota containing low, middle and high education levels are the same in the survey sample and population. The asymptotic significance level is 0,000,

which means the zero hypothesis is rejected because the asymptotic significance level is lower than $\alpha=0,05$ (Table 8-9). The sample is not representative regarding the distribution of the education level in population.

It is possible to weigh the age- and education level classes to balance the non-representativeness of these variables. However, this is not suitable for this thesis research. When classes are underrepresented, it may be risky to weigh these classes such as the non-existing participants who are older than 65 years in the age classes (Table 8-6) or the low educated participants (1,7%) (Table 8-8). The probability that the small groups will represent a bigger group is high. In this case only descriptive statistics are suitable for analysis since inductive analyses are only applicable for random and representative samples (Vocht, 2011). Besides, when the weighing factor is higher than 3,5 it is discouraged to weigh classes (Vocht, 2011).

Table 8-4 Frequencies Gender population in the Netherlands 2016 and survey sample

Gender	The Netherlands		Survey Sample		Weigh factor
	Absolute	Relative	Absolute	Relative	
Man	8417135	49,6	89	49,2	1,0
Female	8561985	50,4	92	50,8	1,0
Total	16979120	100,0	181	100,0	

Table 8-5 Chi-square Goodness-of-fit test Results for Gender

Gender	Test Statistics			
	Observed N	Expected N	Residual	Gender
Man	89	89,8	-0,8	Chi-Square
Female	92	91,2	0,8	df
Total	181			Asymptotic Significance
				0,908
a. 0 cells (0,0%) have expected frequencies less than 5. The minimum expected cell frequency is 89,8.				

Table 8-6 Frequencies age classes in population 2016 and survey sample

Age classes	The Netherlands		Survey Sample		Weigh factor
	Absolute	Relative	Absolute	Relative	
Younger than 20 years	3818499	22,5	31	17,12707182	1,3
20 till 40 years	4163702	24,5	126	69,61325967	0,4
40 till 65 years	5911611	34,8	24	13,25966851	2,6
65 till 80 years	2336560	13,8	0	0	.
80 years and older	748748	4,4	0	0	.
Total	16979120	100,0	181	100	

Table 8-7 Chi-square Goodness-of-test Results for Gender Age classification

Age classification					
	Observed N	Expected N	Residual	Test Statistics	
Younger than 20 years	31	49,8	-18,8		Ageclasses2
20 till 40 years	126	54,3	71,7	Chi-Square	138,170 ^a
40 till 65 years	24	76,9	-52,9	df	2
Total	181			Asymptotic Significance	0,000

a. 0 cells (0,0%) have expected frequencies less than 5.
The minimum expected cell frequency is 49,8.

Table 8-8 Frequencies education level in population 2016 and survey sample

Education level	The Netherlands		Survey Sample		Weigh factor
	Absolute	Relative	Absolute	Relative	
Low	4488000	32,1	3	1,7	19,4
Middle	5405000	38,6	67	37,0	1,0
High	3898000	27,9	111	61,3	0,5
Total	13990000	100	181	100	1,0

Source 1 CBS 2017

Table 8-9 Chi-square Goodness-of-test Results for education level

Education level	Test Statistics			
	Observed N	Expected N	Residual	Education classes
Low education	3	58,9	-55,9	Chi-Square 123,074 ^a
Middle education	67	70,9	-3,9	df 2
High education	111	51,2	59,8	Asymptotic Significance 0,000
Total	181			a. 0 cells (0,0%) have expected frequencies less than 5. The minimum expected cell frequency is 51,2.

Chapter 9 Results

The survey sample is analysed (Chapter 8) and is considered as a selective and non-representative sample. This means only descriptive statistics can apply for the analysis. The analysis scheme presents an overview of the analysis made for the results (Appendix 3). The reliability- and item analysis are also applied for the Likert-scales and multiple response questions (Appendix 4 and 5).

Privacy paradox is often described as dichotomy between behaviour and concerns and attitudes towards privacy and threats (Zafeiropoulou, 2014). In this case, this chapter displays the results from the survey and analyses the dichotomy between the participant's geotagging behaviour on social media and their concerns and attitudes towards location privacy. Attitudes and concerns are not the same, even though both are similar and influence each other. For this reason, both factors will be analysed separately in relation to the participant's willingness to geotag content on social media.

The existence of the privacy paradox can be verified in two ways. The relationship between concern and geotagging behaviour could be examined in two ways: the association between willingness to geotag and concerns. The relationship between participant's attitude and their geotagging behaviour can be examined by studying the association between participant's willingness to geotag their content and attitude positions.

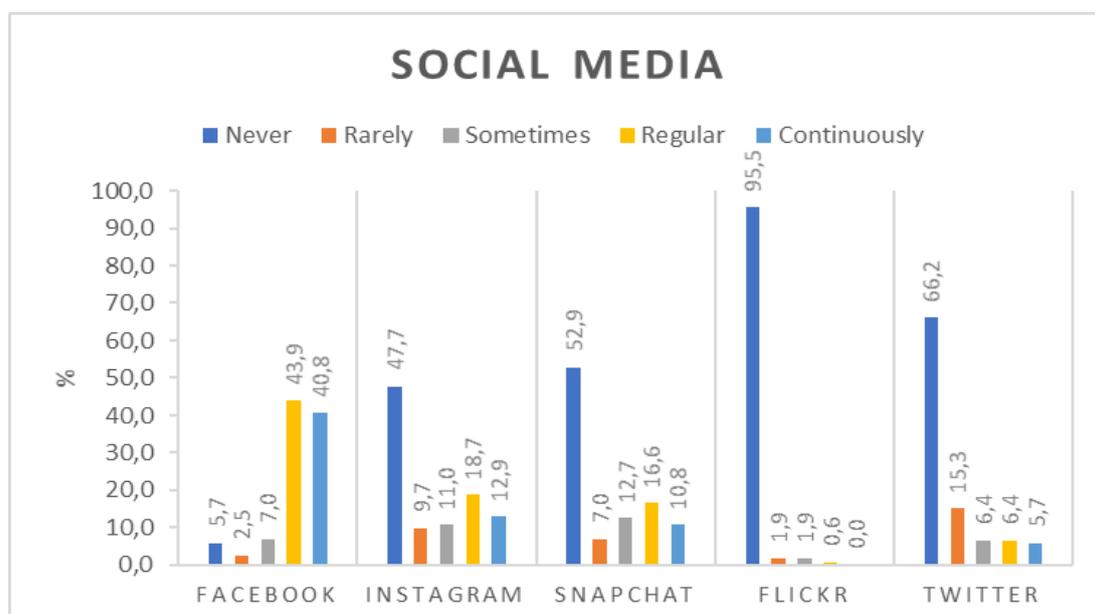
This chapter aims to answer the following sub questions:

- 3) *What is the self-reported geotagging behaviour on social media?*
- 4) *What are the attitudes and concerns of users on location privacy when they are online on their social media profiles?*

9.1 Geotagging behaviour on social media

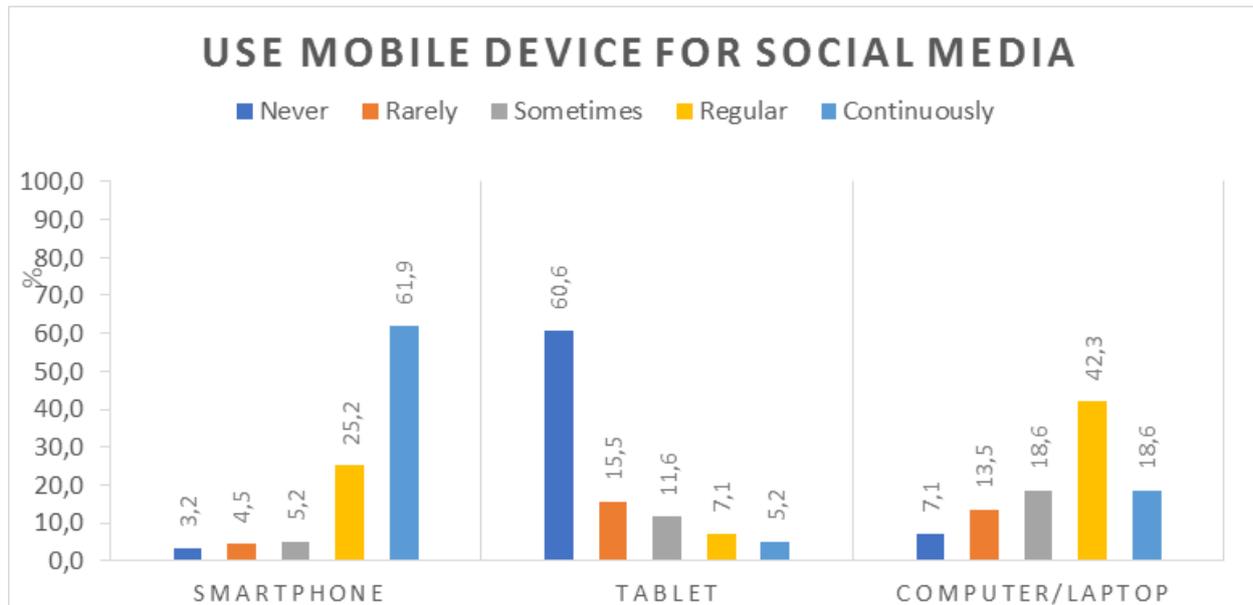
In the beginning of the survey, the participants were asked to indicate which social media platforms they use and at what frequency they use them in their daily lives. Around 155 participants answered the questions about the use of social media use and mobile devices. Around 15% of the participants skipped these questions. The participants use Facebook continuously (40,8%) or regular (43,9%) while participants use Instagram (18,7%) and Snapchat (16,6%) sometimes (Figure 9-1). Twitter (66,2%) and Flickr (95,5%) seem to be least popular among the participants who answered "Never" in

Figure 9-1 Use of social media (N=156)



the survey. Facebook comes to first place of popularity, Instagram second and Snapchat at third place. Smartphone and laptops are the most common among the participants to use for their social media profiles (Figure 9-2). Out of 155 participants 95 are continuously online on social media with their smartphones (61,9%) or regularly on their laptop (42,3%), whereas 95 participants (60,5%) indicate they never use a tablet for social media activities (Figure 9-2).

Figure 9-2 Use of social media on mobile device (N=155)



9.1.1 Geotagging

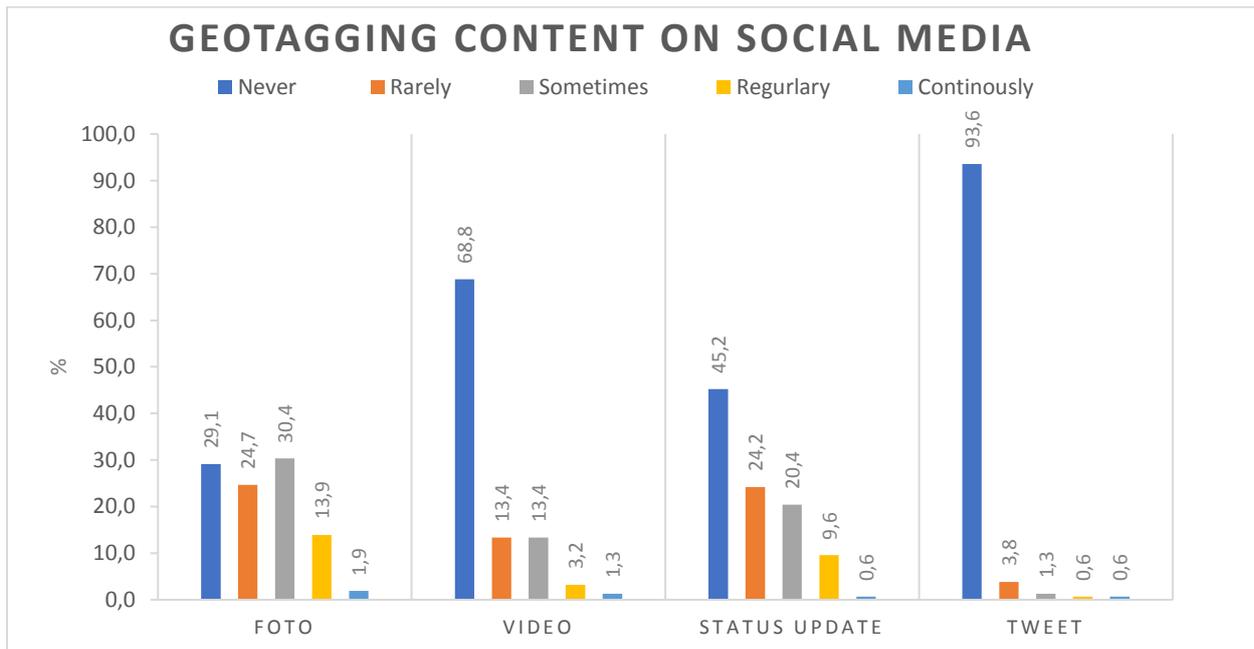
149 participants indicated if they geotag their posts on social media. 35 participants didn't answer the question. 75 of 149 participants (50,3%) use a social media provided list with suggestions to geotag their content, while only 4 participants (2,7%) add places manually with the help of maps (Table 9-1). 56 participants (37,6%) participants claimed they don't geotag their content (Table 9-1). In short, participants (62,4%) geotag their content on social media with the help of a suggestion list also known as *point of interest* and by manually adding location.

Table 9-1 The use of geotag functionality on social media by participants

Response	Frequency	Percent	Cumulative percent
List with suggestions	75	50,3	50,3
Add manually location	4	2,7	53,0
Both	14	9,4	62,4
None	56	37,6	100
Total	149	100	

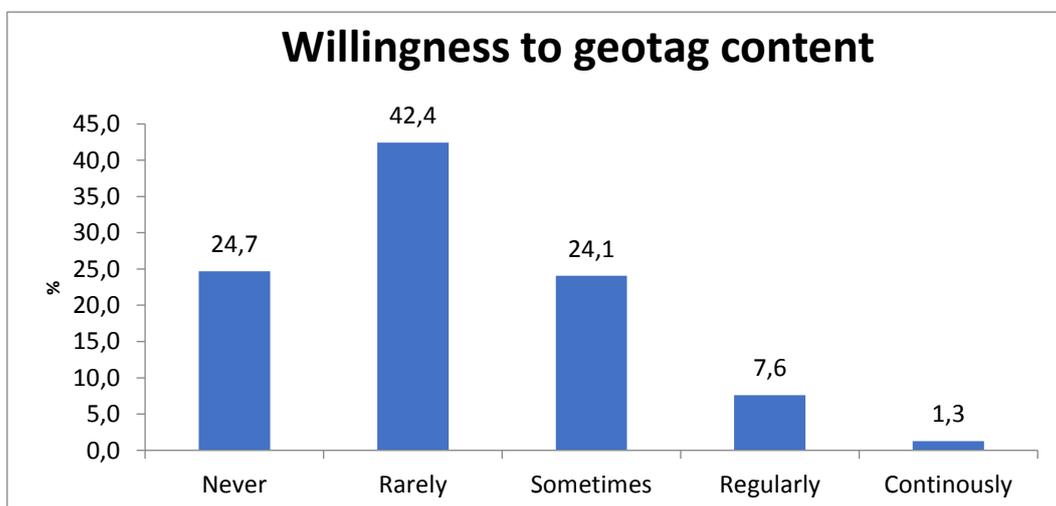
As a follow-up to the geotag feature, participants were asked how often they geotag on purpose in regard to the following content on social media: photo, video, status update and tweets (Figure 9-3). The results show participants don't geotag their content continuously on their profiles. Photos are regularly geotagged (13,9%) and sometimes (30,4%) (Figure 9-3). Besides, participants geotag their status updates sometimes (20,4%), whereas videos (68,8%) and Tweets (93,6%) are never geotagged by most the participants (Figure 9-3). The relatively high number of participants that never geotag Tweets may be explained by the fact 66,2% of the participants never use Twitter.

Figure 9-3 how often do participants geotag their content (N=157)



After the factor and reliability analysis the Likert-Scale of geotagging content is recoded to a new variable "willingness to geotag" (Appendix 4). For each participant, the mean of geotagging content is calculated. 24,7% of the participants never geotag while 42,4% of the participants geotag rarely. The participants who geotag regularly (7,6%) or continuously (1,3%) are relatively small (Figure 9-4). In general, participants rarely geotag their content (Mean= 2, SD=0,94).

Figure 9-4 Willingness to geotag content on social media (N=157)



9.1.2 User characteristics

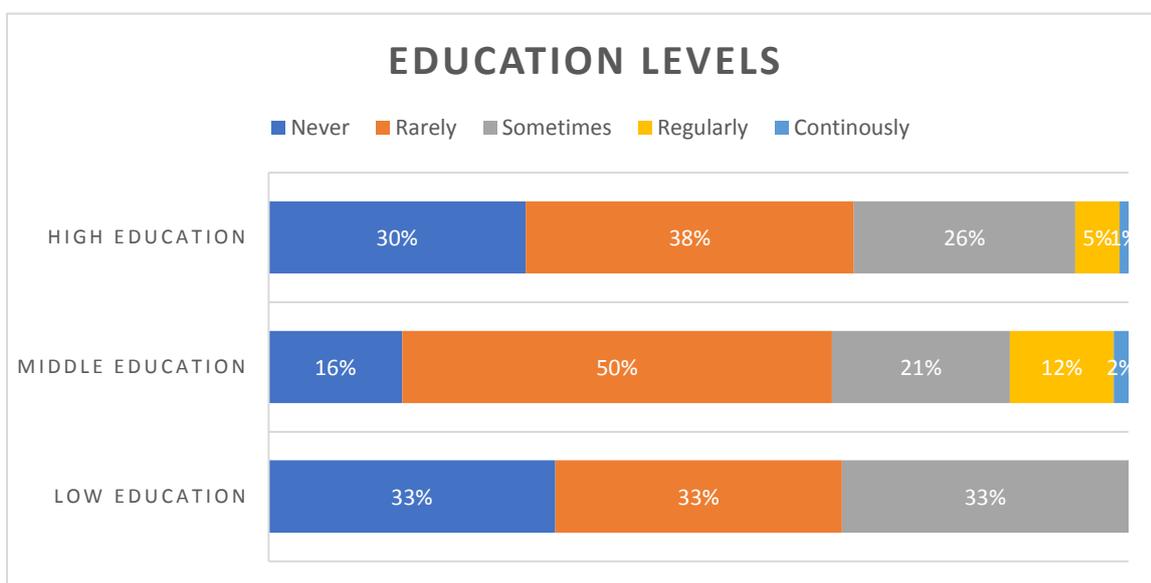
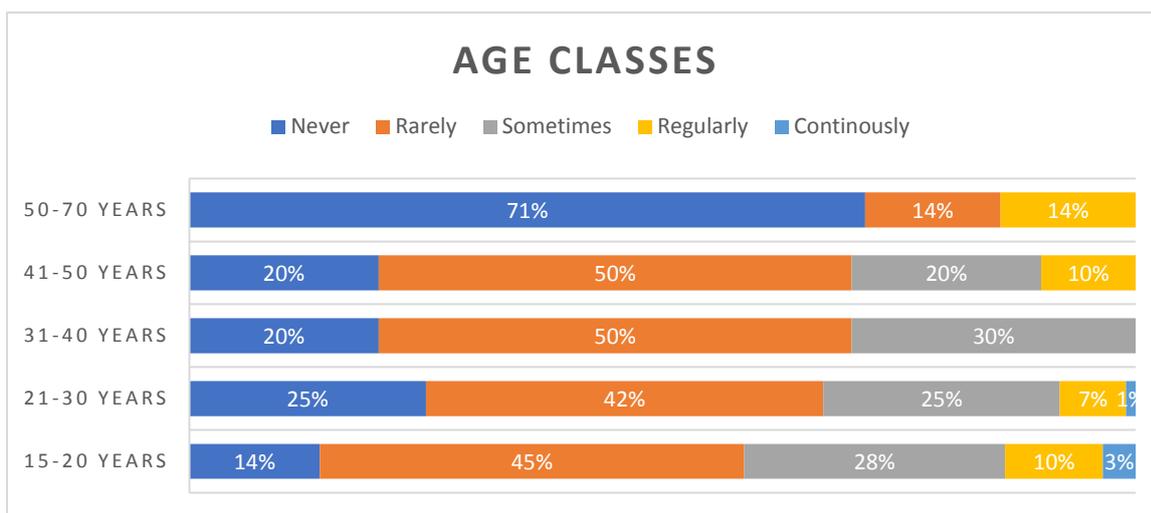
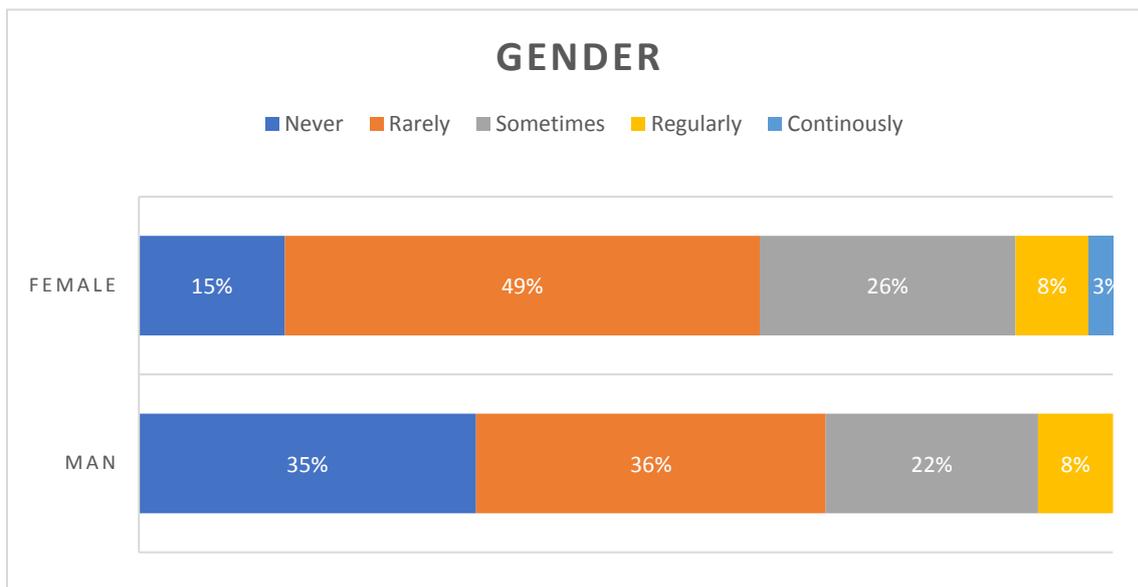
49% of female users rarely geotag content, whereas 36% of man rarely geotag their photos and such (Figure 9-5). Compared to male participants, female participants who geotag their content are represented at a slightly higher amount (26% compared to 22%), whereas the ones who geotag regularly are the same (8%). However, males who never geotag are slightly over-represented (35% compared to 15%). The correlation between gender and willingness to geotag is weak (Cramer's $V=0,251$).

Older participants between 50 and 70 years never geotag (71%), whereas younger participants (15 and 20 years) geotag rarely (45%) and sometimes (28%) (Figure 9-5). The participants who are between 31 years and 50 years rarely geotag content on their profiles. The association between age classes and willingness to geotag is very weak (Kendall's Tau = -0,112).

Highly educated participants (38%) rarely geotag their content, whereas a small group geotag regularly (5%) (Figure 9-5). Middle educated participants geotag rarely (50%), but there is a small difference with the highly-educated participants (38%). According to the Kendall's tau correlation coefficient, the association between education level and willingness to geotag is very weak (Kendall's Tau= -0,088).

To summarise, the demographic characteristics of the participants show a very weak association with their willingness to geotag. However, this does not mean that demographic characteristics influence geotagging behaviour, since the correlation coefficient only refers to association and not to causality between variables or factors of a social behaviour. Since the sample size is not representative for the Dutch population, the geotagging behaviour association with users' characteristics may differ between the sample and population.

Figure 9-5 User's characteristics and willingness to geotag



9.1.3 Places

Participants were asked to indicate the type of places they share on social media. They were free to choose more than one option in the survey. In total 161 out of 184 participants answered this question. Places such as vacation (77%) and trips (25,4%) are viewed as suitable to geotag on social media, whereas workplace (4,9%) or home (5,6%) are less common to geotag on the web (Table 9-2). In question 8 by item "Other", 30 participants also indicated which places they geotag. 17 participants indicated they don't geotag places, whereas 4 participants wrote "trips" as an answer and two participants suggested "restaurants" (Table 9-3).

Table 9-2 Chosen type of places

<i>Response</i>	<i>Frequency</i>	<i>Percent</i>
Vacation	110	57,9
Trips	36	18,9
Home	8	4,2
Workplace	7	3,7
Other	29	15,3
Total	190	100

Table 9-3 Respondent's answers by option: other places

<i>Response</i>	<i>Frequency</i>	<i>Percent</i>
When there is no influence on personal matters	1	3,3
Recreation	1	3,3
Concert	1	3,3
Restaurants	2	6,7
School	1	3,3
Trips	4	13,3
Hobby	2	6,7
Work related	1	3,3
None	17	56,7
Total	30	100

The correlation coefficients Pearson Chi-Square and Cramer's V are calculated to determine if there is an association between the use of geotag feature and types of place (Table 9-4). The findings suggest places like vacation (0.439) and trips (0.341) have a moderate association between geotagging functionality, which is also significant ($\text{sig} = 0,00 < p = 0,05$), though work (0.018) and home (0.159) have very weak association with geotag features (Table 9-4).

Table 9-4 Association between use of geotag feature and type of places

Location	Pearson Chi-Square		Cramer's V
	Value	Asymptotic significance (2-sided)	Value
Vacation	28,8	0	0,439
Trips	17,3	0,000	0,341
Home	3,8	0,052	0,159
Work	0,0	0,826	0,018
Other	7,9	0,005	0,230

9.1.4 Motivations

Location sharing is bound to the motivation of the social media user, which is diverse and mostly socially driven (Tang et al., 2010). 116 out of 184 participants wrote their motivation or reasons why they geotag on social media. Personal reasons motivated participants to geotag their photos and such on their profiles. Participants also liked to share their whereabouts with fellow users on social media and to work on their online self-representation. To some extent the motivations correspond with the expected benefits according to Pötzsch (2009) and Lee et al. (2013) and the social-driven location sharing theory of Tang et al. (2010). The motivations of 116 participants are divided into several different themes. The themes are grouped into the benefits of sharing personal information online according to Pötzsch (2009) and Lee et al. (2013). Each category is divided into sub categories (Table 9-5 and 9-6)⁸. The participants displayed motivations related to relation development (32,8%), social control (16,1%) and the usefulness of the geotag feature on social media (10,9%) (Table 9-5). Usefulness refers to convenience for users to share location information with others without requiring too much effort.

Table 9-5 Expected benefits or motivations to geotag content on social media (N=116)

Benefit	Amount	%
Relation development	45	32,8
Self-presentation	9	6,6
Self-clarification	9	6,6
Social validation	1	0,7
Social control	22	16,1
Usefulness	15	10,9
Fun	12	8,8
Others	3	2,2
No geotagging	21	15,3
Total	137	100,0

⁸ 116 participants answered question 9, which is an open question. In total 9 categories of benefits from geotagging are identified. Table 9-6 presents the sub categories.

Table 9-6 shows the detailed categories with their percentages. To maintain their relations with fellow users on social media, people like to share their location information with others (23,4%) (Table 9-6), while some participants like to geotag their photos and such to give extra information (5,8%) about their whereabouts within a context (3,6%). An interesting outcome is the usefulness of the geotag feature. Some participants mentioned the geotag feature is easy to use (5,1%). However, there were also reasons listed why one does not geotag (12,4%) or share strict location information (2,2%) with fellow users on social media (Table 9-6).

Table 9-6 Detailed categories of benefits geotagging content

Benefit	Categories	Amount	%
Relation development	Everyone does it	1	0,7
	Share with family and friends	10	7,3
	Share with others	32	23,4
	Work	2	1,5
Self-presentation	Interesting places	2	1,5
	Location filter Snapchat	2	1,5
	Self-representation	2	1,5
	Visibility	2	1,5
	Visualization of location information	1	0,7
Self-clarification	Memory	8	5,8
	Reminder	1	0,7
Social validation	To get approval from others	1	0,7
Social control	Context	5	3,6
	Extra information	8	5,8
	Promotion	6	4,4
	Recommendation	2	1,5
	Share knowledge	1	0,7
Usefulness	Easy to use	7	5,1
	To add location	6	4,4
	Useful	2	1,5
Fun	For fun	1	0,7
	Looks nice	11	8,0
Others	Co-location ⁹	1	0,7
	Doesn't know	1	0,7

⁹ Users can post co-location information by tagging friends and families in their posts, thus making location information available to the social media company and fellow users on web.

	Special occasion	1	0,7
No geotagging	Limited sharing due to privacy	3	2,2
	None	17	12,4
	Risks	1	0,7
Total		137	100,0

9.2 Attitude

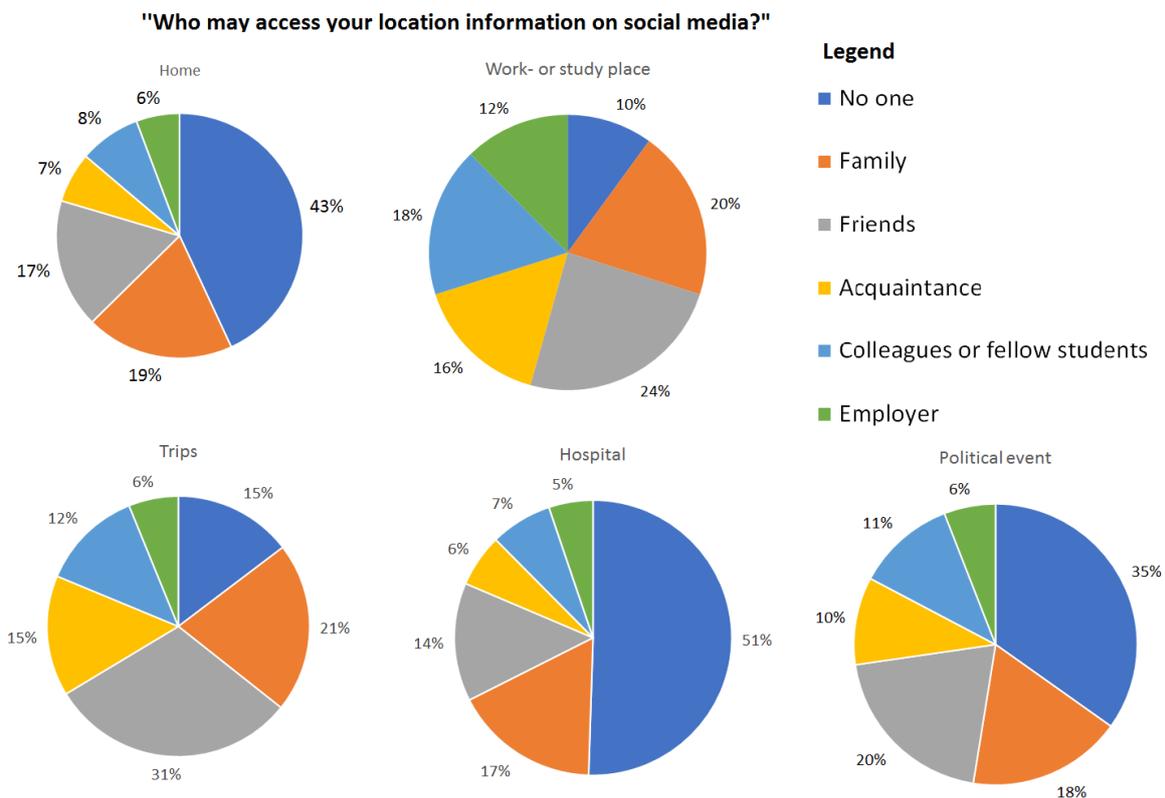
The participants were asked about their attitude on two levels: the permissions that dictate who may see their location information on social media, and which actors may re-use their location information mined from social media. The attitude of the participants is examined towards fellow users on social media who can see their location information. In addition, the participants were asked if they were satisfied with the privacy settings of social media platforms and if they ever changed the settings.

Prior to the results analysis, item analysis has been performed for two latent variable "audience on location information" and "re-use of location information by third parties", which investigates whether the manifest variables measure the latent variable. The manifest variables that measure "audience on location information" are "family", "friends", "acquaintance", "colleague", "employer" and "nobody". The manifest variables that measure "re-use of location information by third parties" are "social media", "companies", "advertisers", "intelligence services" and "research and universities". The item analysis has been elaborated in detail and is listed in Appendix 5.

9.2.1 Audience of location information

The participants were asked to indicate who may see their location information on social media. The following types of places are suggested in the survey: home, workplace, trip, hospital, a political event and no place. The audience is divided into several social relations: family, friends, acquaintance, employer and colleagues, and nobody. Around 135-137 participants (74%) answered these questions which means 26% of 184 participants skipped these questions or refused to answer them. The participants were also permitted to give more than one option as response. So, a participant may opt-in few options for the same type of location.

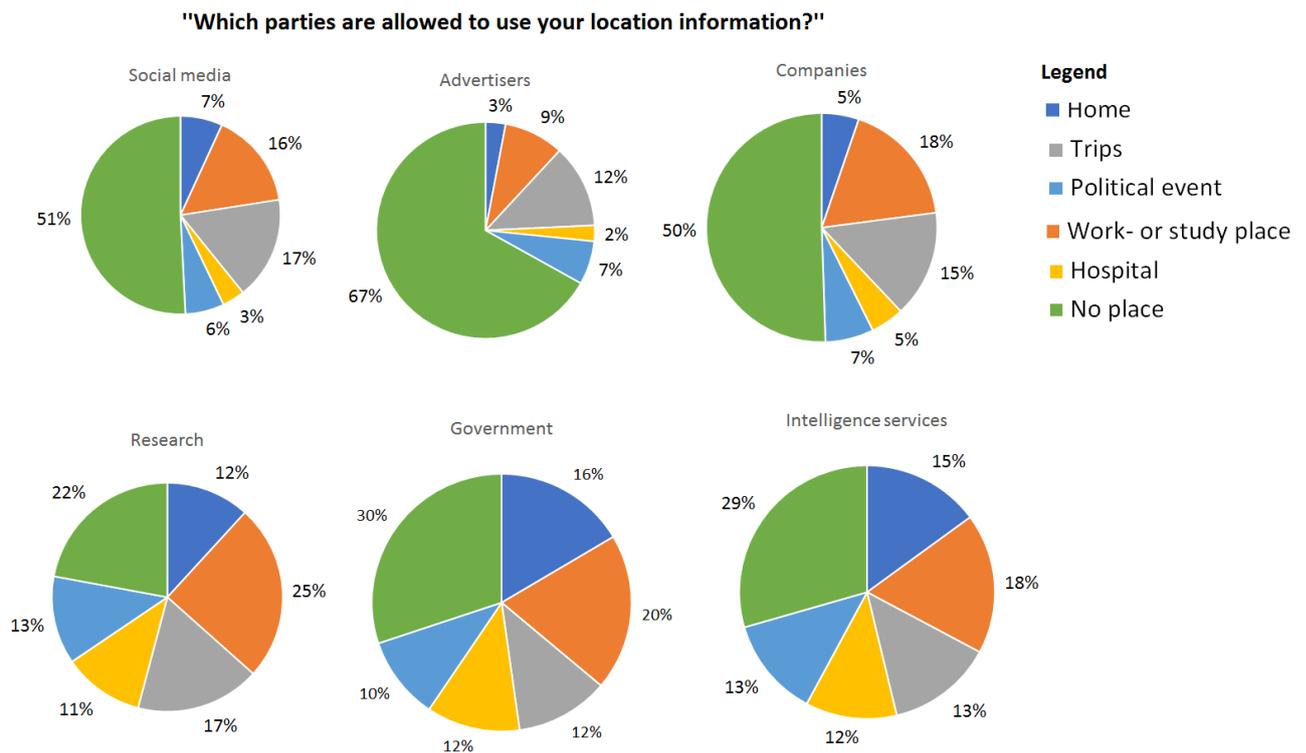
91 participants indicated that they wished nobody could see their home location, and this same sentiment was expressed towards the hospital location (Figure 9-6). Around 50% of the participants wished that hospital stays private on social media. The participants are less strict for others for accessibility to their location information of trips and workplaces. Friends (31%) and families (21%) may view the location of trips. The context, type of relationship and place matters for the participants who might see their location information. For example, employers are less permitted to see location information about trips and hospital compared to other relations such as acquaintances, colleges and friends (Figure 9-6). However, the workplace has a relatively high number of responses towards employers (12%) as a permitted audience compared to trips (6%) or hospital (6%).

Figure 9-6 Attitude towards "audience" on social media ' (N=135-137)¹⁰

9.2.2 Re-use of location data by third parties

The participants were also asked if they permitted access to their location information for re-use by social media companies and third parties (Figure 9-7). The following third parties were suggested in the survey: advertisers, companies, government, intelligence services, research organizations, and universities. The suggested places were again as following: home, workplace, trip, hospital, a political event and no place. In total, 139 participants answered the question related to re-use of location information, whereas 45 participants did not answer. The participants were also permitted to give more than one option as a response. So, a participant may opt-in few options for the same type of location. The results suggest that participants are more critical towards sharing their location information with commercial companies, social media, and advertisers than with parties in research and intelligence services. Furthermore, the results show that, of all commercial parties, respondents are most critical towards sharing their data with advertisers. Over 67% respondents prefer not to share any location information with advertisers.

¹⁰ The question "Who may have access to your location information?" is the free translation and summarized version of the statements of question 12.

Figure 9-7 Attitude towards re-use of location information by third parties (N=139)¹¹

9.2.3 Privacy attitudes: interest positions

The participants were asked to answer which description of attitude matches best with their attitude: fundamentalist, pragmatist and unconcerned. 135 participants indicated their attitude towards privacy on social media, whereas 49 participants (26%) did not answer this question (Table 9-7). Most participants (58,5%) identified themselves as privacy pragmatists, while 27 participants (27, 4%) called themselves privacy fundamentalist. A small group of 4 participants (3%) did not identify themselves with any of the attitude descriptions as suggested in the survey. The participants who are privacy unconcerned represent a small group of 15 persons, making them 11, 1% of the valid cases.

Table 9-7 Participant's attitude towards online privacy

Attitude towards privacy	Frequency	Valid Percent	Cumulative Percent
Fundamentalist	37	27,4	27,4
Pragmatists	79	58,5	85,9
Unconcerned	15	11,1	97
None	4	3	100
Total Valid	135	100	
Missing	49		
Total	184		

¹¹ The question 'Which parties are allowed to use your location information?' is the free translation and summarized version of the statements of question 13.

9.2.4 Dichotomy between privacy interest position and willingness to geotag

The association between the participants' willingness to geotag and their self-reported attitude is moderately strong (Cramer's $V=0,354$). 58% of participants ($N=135$) identified themselves as privacy pragmatist, while a smaller group identified themselves as privacy fundamentalist (27%) (Table 9-7). 56% of the participants who never geotag considered themselves to be privacy fundamentalist, whereas 24% consider themselves to be privacy pragmatist (Table 9-8). The participants who rarely geotag, which is also the biggest group, exists of privacy pragmatist (59%), privacy fundamentalist (29%) and a small group of persons consider themselves as unconcerned (3%).

Table 9-8 Cross table between privacy interest position and willingness to geotag

Privacy Interest position	Willingness to geotag										Total
	Never		Rarely		Sometimes		Regularly		Continuously		
	N	%	N	%	N	%	N	%	N	%	
Fundamentalist	18	56	17	29	1	3	1	9	0	0	37
Pragmatist	11	34	34	59	26	81	7	64	1	50	79
Unconcerned	2	6	5	9	4	13	3	27	1	50	15
None	1	3	2	3	1	3	0	0	0	0	4
Total	32	10	58	100	32	100	11	100	2	10	135
		0								0	

9.3 Concerns

The participants were asked about their concerns related to location privacy on two levels: the threats on their privacy and misuse of their location information by several parties such as advertisers or government. Prior to the results analysis, reliability analysis has been performed for the latent variable "location privacy concern", which investigates whether the manifest variables measure the latent variable. The reliability analysis has been elaborated in detail and is listed in Appendix 4.

9.3.1 Concerns regarding location privacy

The concerns regarding location privacy are also examined via Likert-scales with 5 items. The participants answered the statements within 5 levels. Most participants are concerned about their location privacy as related to companies and advertisers. However, the participants are less concerned about the risks related to identity fraud or social issues as a result of sharing location information on social media. Their concerns regarding the misuse of location privacy by universities and research institutions are under-represented among the participants (Figure 9-9). 59 % of the participants indicate they do not agree with the statement about the concerns related to universities. However, participants are generally quite neutral about privacy threats regarding location privacy ($M=3,12$, $SD=0,94$), but also towards re-use of their location information by third parties such as companies and research institutions ($M=3$, $SD=0,98$). Both components are measured on scale 1 till 5¹².

Most participants are not concerned about privacy threats such as identity theft (55.1%), social issues on their social network (67,4%) or invasion of privacy by hackers (51%). Nor are they concerned about the re-use of location information by third parties such as the government (71,5%) and companies (57,1%). Although, participants (70%) are concerned about the re-use of location information by advertisers.

¹² See Appendix 4 for factor- and reliability analysis on Likert-Scale Concerns. The mean of each participant and therefore the mean of the Likert-scale are calculated on two components "privacy threats" and "re-use by third parties".

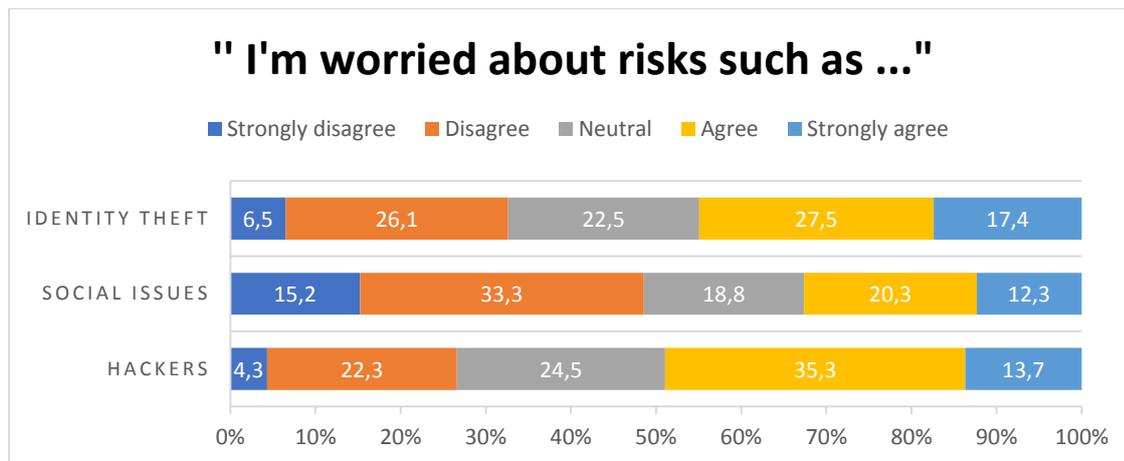
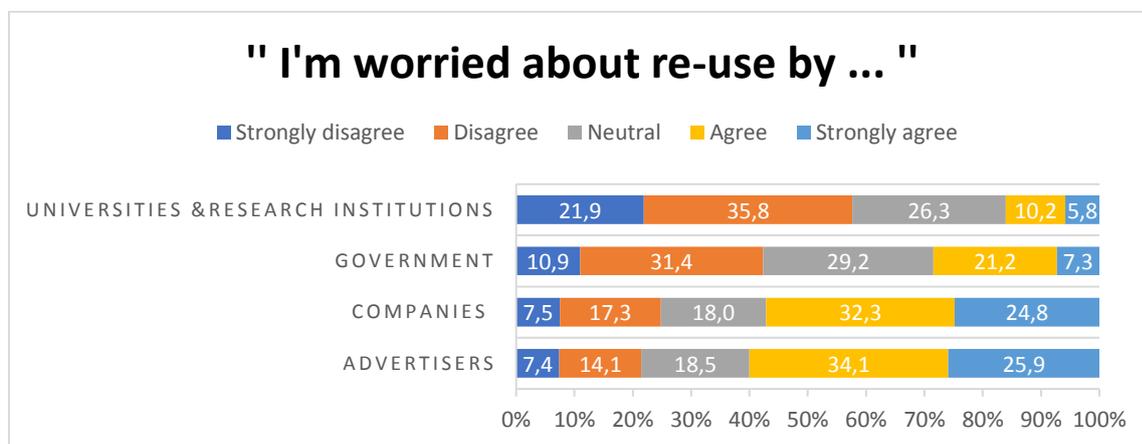
Figure 9-8 Concerns regarding location privacy and risks¹³ (N=136)

Figure 9-9 Concerns regarding location privacy and actors (N=136)



9.3.2 Dichotomy between concerns and willingness to geotag

In general, participants are neutral towards privacy risks and threats by re-use of location information. Participants rarely geotag their content and only for special events such as vacation or trips. Nevertheless, there is a negative weak correlation (Spearman's $Rho = -0,078$) between the concerns regarding privacy threats and the willingness of the participants to geotag, which is also the same for the correlation between concerns related to re-use by third parties (Spearman's $Rho = -0,212$). Because participants are generally not very concerned and do not geotag regularly, it is hard to say whether there is a privacy paradox for geotagging when one considers participants concerns and geotagging behaviour as a dichotomy.

¹³ The titles "I'm worried about..." is the free translation and summarized version of the statements of question 15.

Chapter 10 Conclusion

This chapter aims to answer the research questions that were posed in the second chapter. The objective of this research is to investigate whether privacy paradox applies to location information by studying the user's privacy concerns on, and attitude towards, location privacy, and user geotagging behaviour on social media. By making use of quantitative analysis and literature and desktop research, the privacy paradox on geotagging has been studied.

The first objective is achieved with the help of desktop and literature studies in chapters 4, 5 and 6. The second and third objectives are achieved with the help of data-analysis performed on the results of the online survey, which was filled in by 184 social media users. To answer the main research questions, the sub research questions will be answered first.

1) *To what extent is location information considered to be (sensitive) personal data by law?*

a) *What is location information?*

Location information is defined as data with geographic coordinates of a person or object that is identified by positioning technologies within certain accuracy (Section 4.1). The thesis follows the definition of location information as it is defined by E-Privacy Directive 2002/58/EC: "(14) Location data may refer to the latitude, longitude and altitude of the user's terminal equipment, to the direction of travel, to the level of accuracy of the location information, to the identification of the network cell in which the terminal equipment is located at a certain point in time and to the time the location information was recorded." Location data is processed in an electronic communications network and indicate geographic position of the terminal equipment of a user in a publicly available service (Directive 2002/58/EC article 2).

Privacy is also a context-dependent concept due its integrity. That means the norms of privacy may differ for each person and group, but also depend upon who is on the receiving end of the information flow. The type of disclosed information and the use to which information will be put also influences one's perception on privacy. Location information may enable identification of individuals' interest on politics, religion or other themes. This also touches upon location privacy of users. Location privacy, a special type of informational privacy, is the right of an individual to protect his location information from disclosure or to determine the extent to which data can be shared (Section 4.2).

b) *Under which circumstances is location data considered as personal information?*

The context of the information determines its sensitiveness (Section 4.3). According to General Data Protection Regulation (GDPR), location information is considered as personal data, as information that relates to an identified or identifiable natural person. Location data might contain (sensitive) personal information that refers to a physical, physiological, genetic, economic or social identity of a natural person. When it does it is considered to be sensitive information according to the European Data Protection Directive 95/46/EC.

2) *What is geotagging?*

a) *How do geotag features work on social media?*

Nowadays social media users act as social sensors with the help of their mobile devices and create content such as photos with geographic coordinates (Section 5.1). This process is also known as geotagging. The created content contains personal information about the identified mobile device and gives context and information to one's movement, behaviour, and interest.

Social media users geotag their digital content such as photos, videos and text messages on social media. One can directly choose a place name from a list with suggestions (point of interest) or manually add places. Most social media platforms are embedded with geotag features. The geotag functionality requires the involvement of one of the technologies that position the location of a user's mobile

device: cellular identification, GPS, Wi-Fi hotspots, assisted GPS and Internet Protocol Address (IP) (Section 5.2). Every method has its own positioning technique and an accuracy of measurement.

b) *For what reasons do social media companies collect personal data?*

When users geotag their content on social platforms they do not only produce location data that is associated with the content, but other personal data such as metadata, cookies and beacons, which also contain location information about the used mobile device (Section 6.1 and 6.2). Social media companies collect this personal data for several reasons such the improvement and personalization of their services, personalization with advertisements, and to improve their recommendation mechanism on social platforms. Social media companies also resell personal data to third parties for online advertising and customer relationship management. The personal data is gathered and repackaged in a way that is relevant for advertisers to buy and re-use it for identifying potential buyers via behavioural targeting based on users' characteristics such as age, gender and location information.

c) *What are the risks of geotagging regarding the privacy of social media users?*

When a dataset with location information and other personal data are combined, this might significantly increase the risk of privacy threats for users (Section 6.3). Location information can easily be obtained based on users' check-ins on places via social media and geotagging of content on web. With the help of location data, the daily movement of individuals and behaviour can be predicted and analysed. Users may become victim of threats such as surveillance, identity theft or cyber stalking (Section 6.3).

The risks of geotagging or sharing location information are grouped into 4 dimensions that may affect location privacy of an individual. First, the amount of collected data and its quality defines the level of detailed profiling. Second, the type of location data (spatial semantic, non-spatial semantic and temporal semantic) provides the accuracy of the location information. Third, the accessibility to data with location information refers how much of the user's data is available and visible to others. Social media users have limited access to their own collected data. However, APIs provide access to full publicly available data of users. This accessibility also creates the possibility for exploitation of location information by third parties. Fourth, lack of security and data protection may also stimulate misuse of personal data. Therefore, developers of location-based social platforms should consider the risks of privacy threats and provide security to protect data and improve privacy policy related to re-use of location information by third parties and other consequences. Despite the privacy risks, users keep using social media and share location information on social media with their fellow users.

3) *What is the self-reported geotagging behaviour on social media?*

Participants' self-reported geotagging behaviour on social media is explored via an online survey. The geotagging behaviour is defined by specific user activities on social media: (1) use of social media, (2) frequency of geotagging content and (3) motivation to geotag. The participants rarely geotag content on their profiles. Photos and text are most frequently geotagged, whereas Tweets or videos are never or rarely geotagged. Facebook and Instagram are most popular among participants, whereas Twitter and Flickr are least popular. According to the correlation coefficients there is weak association between the users' characteristics and their willingness to geotag. Neither gender, age, nor education level has a strong association with the participants' willingness to geotag content on social platforms. Participants were also asked why they geotag. The reasons touch upon the benefits users expect from sharing location information. Common benefits are related to relation development (share with friends and families), social control (to create context and give extra information to fellow users), and usefulness (ease of use). Some participants also claimed they do not geotag due caution with (over) sharing personal information on web.

4) *What are the attitudes and concerns of users on location privacy when they are online on their social media profiles?*

The participants' concerns regarding privacy threats are examined by re-use of location data by third parties and other kind of risks such as identity fraud. In comparison to other scientific works on privacy concerns which claimed that users are concerned about their online privacy, these participants are neutral about it. Participants are generally neutral towards privacy risks by threats such as identity fraud or by re-use of location information by third parties such as government or companies (Section 9.3). The findings in Section 9.3 show that users are not fully aware of the threats and risks related to their location privacy. Most participants are not concerned about privacy threats such as identity theft (55.1%), social issues in social networks (67,4%) or invasion of privacy by hackers (51%), nor are they concerned about the re-use of location information by third parties such as government (71,5%) and companies (57,1%). There is also a negative weak correlation between the participants' concerns and their willingness to geotag content on social platforms. There is no privacy paradox on geotagging behaviour when one considers the neutrality of the participants towards the privacy risks.

In addition, the attitude of participants is also examined by looking at the self-claimed privacy interest position, attitude towards re-use by third parties and permission to access their personal location information (Section 9.2.1 and 9.2.2). Participants are cautious about sharing location information with other users and third parties. Usually, participants do not want to share their location information with fellow users, nor do they want to share their location information with advertisers and companies. Participants distinguish between types of relationships at a type of place. For example, according to participants, the employer should not be able to see where they were on holiday, but the employer is, however, permitted to see the participants' workplace. This finding is consistent with Nissenbaum's theory (Section 4.2). Nissenbaum (2011) suggests that one's understanding of privacy is dependent on the context of an activity and on one's expectation of which information should be private or public.

Participants identified themselves as privacy pragmatist or privacy fundamentalist and geotag their photos and text posts on their profiles rarely or only sometimes (Section 9.2.3). The ones who never engaged in geotagging were overrepresented by privacy fundamentalist, whereas the ones who geotag rarely and sometimes were overrepresented by pragmatists. The self-claimed privacy interest position (attitude) has shown a moderately strong association with the willingness to geotag of participants. There is no privacy paradox between the participants' attitude and their geotagging behaviour.

The main research question of this research can now be answered:

Does the privacy paradox theory apply to geotagging on social media, and if so how may the privacy paradox be overcome?

This thesis suggests that the privacy paradox does not apply to geotag behaviour on social media since participants geotag rarely, and when they do, it is for special occasions. Additionally, participants are neutral about the privacy concerns. However, participants are also cautious towards sharing location information with fellow users on social media. The type of relation on social media and the type of place affect the decision-making of users deciding who is permitted to see their whereabouts. Moreover, type of actor (government versus non-government) also matters for the participants in terms of re-using location data for several purposes. Participants provide governmental parties more access to location information compared to non-governmental parties. These findings point out that privacy is indeed a contextual-dependent concept and is also connected to individual's perception of location privacy.

Despite the conclusion of this thesis, there is still a chance that privacy paradox exists regarding location information, based on the academic literature on online behaviour and privacy concerns of users. Social media users rarely geotag their content, but social media companies can still identify the location of mobile devices via positioning technologies such as GPS and Wi-Fi triangulation. Furthermore, social media companies can obtain location information from metadata of created

content and by tracking cookies on web. The mined personal data of the users becomes an asset of social media companies and advertisers for profitable objectives.

The results show that users consider whether to geotag their content or not. Still, they do not have control over their own personal data flow, which is a violation of an individual's right to privacy. Users are also not fully aware that social media platforms collect many different types of location information, and share this data for several purposes with various third parties. For example, location information is distributed to location-based marketing companies to analyse users' behaviour and to advertise with personalized advertising on social platforms. The information flow of the location data between the user, social media and other parties is unknown to the users. Therefore, it is hard for users to determine whether it is prudent to share location information on web.

This informational asymmetry between users and social media companies should be resolved by supporting the user in identifying and assessing privacy risks from disclosing location information on web. It is necessary to create more privacy awareness among users by solving the information asymmetry. The gap can be closed when more information is available on the following subjects:

- Collection of location data by social media companies;
- Storage of location data;
- Re-use of location data by third parties;
 - Identification of third parties.

Social media companies such as Facebook and Instagram need to be more transparent towards their users about the monetization of location data. Governments and interest groups for digital citizen rights or privacy should design toolkits and provide guidelines based on the information about personal data flow in social media. The starting point of the toolkits and guidelines should follow the assumption that privacy is a context-dependent, dynamic concept. Policy makers and developers are advised to consider the flexible boundary between private and public sphere in the real and digital world.

Chapter 11 Discussion

In this chapter, a critical reflection will be provided by evaluating the results and the applied methodology (Section 11.1). Thereafter, recommendations for further research will be provided based on the conclusion and the reflection (Section 11.2). In addition, suggestions for creating more privacy-awareness around risks of location information sharing on applications will be provided (Section 11.2). The suggestions are made to achieve the third objective of this research.

Privacy paradox may not be the only outcome of an unconcerned attitude towards privacy risks, but rather because of a lack of awareness with regards to the disclosure behaviour on personal – and location information and its possible consequences (Deuker, 2009). As mentioned before, informational asymmetry exists between services and social media users about data storage, and usage and dissemination of personal information to third parties. Regarding the informational asymmetry, users need to be supported to be able to identify and assess risks associated to the disclosure of personal and location information (Deuker, 2009; Pötzsch, 2009). Creating more awareness among users may solve the privacy paradox (Mascetti et al., 2011).

11.1 Reflection

The expected outcome of this study is expressed with the hypothesis: **The privacy paradox does exist on geotagging behaviour on social media.** The hypothesis was based on the theory of the dichotomy between the concerns of users related to privacy and their disclosure behaviour. Social media users trade their location information to gain benefits and to use "free" services. Meanwhile, this is paradoxical in comparison with their highly stated concerns and attitudes (Zafeiropolou, 2014). However, the findings of this study show that there is no privacy paradox occurring regarding geotagging behaviour on social media. According to the theory, users tend to have a relatively high concern about their privacy online. On the other hand, users reveal a lot of personal information on social networks for relatively small rewards and the attention of peers (Kokolakis, 2015). However, the findings show that participants are generally neutral towards privacy threats. As for their attitude towards location privacy, in general, participants are cautious with providing access to their location information to third parties and to fellow users on social media. This finding corresponds well with the fact that participants rarely geotag their content on social media such as Facebook and Instagram. In other words, the findings of the thesis do not completely correspond with the formulated hypothesis and the general theory on privacy paradox.

The difference between the conclusion of this thesis and theory on privacy paradox might be due the operationalization of disclosure location information. Zafeiropolou (2014) and Furini & Tamanini (2014) operationalized sharing location information as an activity wherein users give the authorization to applications and services to save location data of their mobile devices. In this research, the focus is on geotagging behaviour of users wherein an individual consciously shares their location information in combination with digital content such as a photo or video. Both behaviours are not the same, and this might lead to different findings related to privacy paradox on location information.

Furthermore, the methodology relies on the self-reported behaviour of the participants. Partially, the validity of the results depends on the participants' collaboration to fulfil the survey seriously. For further research, it is recommended to analyse data mined from social media such as Instagram, Twitter or Snapchat. The user's characteristics, choice of geotagged content and places that are geotagged can be studied effectively via social media data. This information will provide more knowledge on the geotagging behaviour of users.

As mentioned before in chapter 3 and 8, participants are non-randomly selected with the snowball method and the survey sample is not representative for the Dutch population. Young and highly-educated users are over-represented in the sample. This touches upon one of the remarks Sloan and Morgan (2015) made in their studies on Twitter users. Users who geotag hardly represent the overall population, since a relatively small group of users geotag their content on social platforms. Just

in case, if a representative sample is highly required for a research it would be better to work with random selective methods for surveying. Besides, the sample size is relatively small to conclude that privacy paradox does not apply to young and highly educated users. Studies with representative samples are necessary to generate an overall conclusion related to social media users. An alternative solution may be a cooperation with research agencies to work with non-random sampling methods and reach many as possible respondents for research.

The online survey is a strategic choice that allows the researcher to reach many participants in a very short time. However, it seems that filling in the survey might take too long for some participants. The non-response rate was relatively the highest at the final questions regarding attitude towards location privacy and concerns. Additionally, the design of the survey should be more optimized for data-analysis. The question regarding attitude towards audience on social media (Q12) and re-use of location information (Q13) results in too many categorical variables, which made it hard to analyse the dichotomy between the participant's willingness to geotag and their attitude towards location privacy. Because of this, only the privacy interest positions of the participants (Q14) were used to determine the strength of the correlation between user's willingness to geotag and their attitude towards location privacy.

Every research has its limitations and therefore it is important to be critical towards methodology and the presented results. Nevertheless, this research contributes new insights to the perception of location privacy by social media users and sets up new suggestions to stimulate more privacy-awareness among users.

11.2 Recommendation

Recommendations are provided based on given conclusion and discussion. These recommendations may lead to further research on geotagging behaviour, location privacy and usage of location data generated from social media. Despite relatively low geotagging by users, the usage of location-based services and social platforms will eventually grow in coming years. There is a need for more knowledge on whether users decide to geotag their content on social and commercial platforms. For this reason, the following research topics are recommended:

1. Type of places

Exploratory research on geotagging on social platforms will provide more insight on users' geotagging behaviour and their privacy trade-off on social media. The location of users' activity influences the willingness to geotag content on social media. Fun activities such as going on holiday, dinner or going out are more likely to be geotagged for socially driven purposes such as self-representation and relation development. Workplaces and home addresses are less likely to be geotagged by users.

2. Privacy management

Another approach as a research topic would be the presence of fellow users and their influence during the privacy trade-off on social media. Users on social media are cautious regarding the audience on their profiles and the accessibility to their location information, and other kinds of personal information. The kind of relation with fellow users and the context of the content determine the willingness to geotag. Family and friends are considered as trustworthy most of the time, whereas users are more cautious towards employers, co-workers and acquaintances.

Another approach to privacy management is the focus on usage of privacy policies in social media and the privacy settings of applications. Applications on mobile devices ask explicitly if the application may have access to data that is stored on the device, but also asks permission if the application may identify the location of the mobile device. Research on privacy management would be useful to create more privacy awareness among mobile device users.

3. Motivations and expected benefits from geotagging

Users share personal information and location information to gain expected benefits from the applications and its services which they use. The motivation to share location information with location-based applications is studied several times. However, a qualitative research on users' motivations will provide more in-depth knowledge about users' willingness to geotag their content and their motivations to do so. Qualitative research is a suitable method to explore motivations and perception on social phenomena.

4. User interface design

As mentioned before in the theoretical chapter, social media companies use interface design to manipulate their users into sharing personal information in many ways by asking questions such as "What is on your mind?", or by enabling the functionality to post content on different social platforms. Instagram is also embedded into Twitter and Facebook, which means that users can post their Instagram photos at the same time on Twitter and Facebook as well. The connection between social media networks and its usefulness stimulates users to share more photos with fellow users. Studies with more focus on user interface design would be refreshing and interesting for research on privacy trade-off on social media.

5. Focus on different social media

Facebook and Twitter are framed as the older generation of social media, whereas Instagram and Snapchat have grown in number of young users. Most studies were focused on privacy trade-off on Facebook, while Twitter and Flickr are used to mine data for studies based on data-analysis. Therefore, research on the "upcoming" social media platforms such as Snapchat would be interesting, since Snapchat also stimulates her users to use GeoFilter¹⁴ for their photos.

¹⁴ Geofilters are location based overlays that users can apply to their photos and videos.

References

- Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, 347(6221), 509-514.
- Acquisti, A., & Gross, R. (2006). Imagined communities: awareness, information sharing, and privacy on the Facebook. *International Workshop on Privacy Enhancing Technologies*, Springer, Berlin, 36–58. https://doi.org/10.1007/11957454_3
- Alrayes, F. S., & Abdelmoty, A. I. (2014). No Place to Hide: A Study of Privacy Concerns due to Location Sharing on Geo-Social Networks. *International Journal on Advances in Security*, 7(34), 62–75. Retrieved from <http://orca.cf.ac.uk/70577/>
- Armstrong, M. P., & Ruggles, A. J. (2005). Geographic information technologies and personal privacy. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 40(4), 63-73.
- Autoriteitpersoonsgegevens.nl. (2017). Dutch data protection authority: Facebook violates privacy law. [online] Available at: <https://autoriteitpersoonsgegevens.nl/en/news/dutch-data-protection-authority-facebook-violates-privacy-law> [Accessed 5 June 2017].
- Beresford, A. R., & Stajano, F. (2003). Location privacy in pervasive computing. *IEEE Pervasive Computing*, 2(1), 46–55. <http://doi.org/10.1109/MPRV.2003.1186725>
- Barkhuus, L., & Dey, A. (2003). Location-Based Services for Mobile Telephony: a Study of Users' Privacy Concerns, (C), 709–712.
- Barkhuus, L., Brown, B., Bell, M., Hall, M., Sherwood, S., & Chalmers, M. (2008). From awareness to repartee: Sharing Location within Social Groups. *Proceeding of the Twenty-Sixth Annual CHI Conference on Human Factors in Computing Systems - CHI '08*, 497. <https://doi.org/10.1145/1357054.1357134>
- Bergström, A. (2015). Online privacy concerns: A broad approach to understanding the concerns of different groups for different uses. *Computers in Human Behaviour*, 53, 419–426. <https://doi.org/10.1016/j.chb.2015.07.025>
- Benisch, M., Kelley, P. G., Sadeh, N., & Cranor, L. F. (2011). Capturing location-privacy preferences: quantifying accuracy and user-burden tradeoffs. *Personal and Ubiquitous Computing*, 15(7), 679-694.
- Blumberg, A. J., & Eckersley, P. (2009). On Locational Privacy, and How to Avoid Losing it Forever By On Locational Privacy, and How to Avoid Losing it Forever. *Electronic Frontier Foundation Tech Rep August*, 1–7. <http://doi.org/10.1109/icc.2011.5962528>
- Boeije, H., Hart, H., & Hox, J. (2009). *Onderzoeksmethoden* (1st ed.). Den Haag: Boom Onderwijs.
- Bryman, A. (2012). *Social research methods*. Oxford: Oxford University Press.
- Chang, J., & Sun, E. (2011). Location 3: How Users Share and Respond to Location-Based Data on Social Networking Sites. In *Proceedings of the Fifth AAAI Conference on Weblogs and Social Media* (pp. 74–80). <https://doi.org/papers3://publication/uuid/3EE04FFA-96B9-4B39-81EE-B42C6BC39F33>
- Chin, A., & Zhang, D. (Eds.). (2014). *Mobile social networking*. Springer. Retrieved from <http://www.springer.com/series/11784>

- Cho, H., Lee, J. S., & Chung, S. (2010). Optimistic bias about online privacy risks: Testing the moderating effects of perceived controllability and prior experience. *Computers in Human Behaviour*, 26(5), 987–995.
<https://doi.org/10.1016/j.chb.2010.02.012>
- Cramer, H., Rost, M. & Holmquist, L.E., 2011. Performing a Check-in: Emerging Practices, Norms and “Conflicts” in Location-Sharing Using Foursquare. *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, 57–66.
Retrieved from: <http://dl.acm.org/citation.cfm?id=2037384>
- Christin, D., Reinhardt, A., Kanhere, S. S., & Hollick, M. (2011). A survey on privacy in mobile participatory sensing applications. *Journal of Systems and Software*, 84(11), 1928-1946.
- Conti, G., Point, W., & York, N. (2010). Malicious Interface Design: Exploiting the User, 271–280.
- Cottrill, C. D., & Thakuriah, P. V. (2012). Consumer location privacy preferences: Survey analysis. In *Transportation Research Board 91st Annual Meeting* (NO 12-4731).
- Cvrcek, D., Kumpost, M., Matyas, V., & Danezis, G. (2006, October). A study on the value of location privacy. In *Proceedings of the 5th ACM Workshop on Privacy in Electronic Society* (pp. 109-118). ACM.
- Danezis, G., Lewis, S., & Anderson, R. (2005). How much is location privacy worth? *Fourth Workshop on the Economics of Information Security*, 78, 739–748. <https://doi.org/10.1111/j.1747-0285.2011.01230.x>
- de Hert, P., & Papakonstantinou, V. (2016). The new General Data Protection Regulation: Still a sound system for the protection of individuals? *Computer Law & Security Review*, 32(2), 179-194.
- Deuker, A. (2009). Addressing the privacy paradox by expanded privacy awareness – The example of context-aware services. *Privacy and Identity Management for Life in IFIP Advances in Information and Communication Technology*, 320, 275–283.
- Dourish, P. (2006). Re-space-ing place: place and space ten years on. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work* (pp. 299-308). ACM.
- Dinev, T. (2014). Why would we care about privacy? *European Journal of Information Systems*, 23(2), 97–102. <https://doi.org/10.1057/ejis.2014.1>
- Duckham, M., & Kulik, L. (2006). Location privacy and location-aware computing. *Dynamic & mobile GIS: investigating change in space and time*, 3, 35-51.
- Elwood, S., & Leszczynski, A. (2011). Privacy, reconsidered: New representations, data practices, and the geoweb. *Geoforum*, 42(1), 6–15. <https://doi.org/10.1016/j.geoforum.2010.08.003>
- *europe-v-facebook.org* | *EUROPE versus FACEBOOK*. (2017). *Europe-v-facebook.org*. Retrieved from <http://europe-v-facebook.org/EN/en.html> [Accessed 3 March 2017]
- European Parliament and Council (1995). Directive 95/46/EC of the European Parliament and of the council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal L 281*, 0031 - 0050
- European Parliament and Council (2002). Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications). *Official Journal L*, 201, 0037 – 0047

- European Parliament and Council (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *OJ L 119*, 1–88
- Hay, I. (2010). *Qualitative research methods in human geography*. Oxford: Oxford University Press.
- *How accurate is IP Geolocation?* (2017). *What is my IP address*. Retrieved 1 June 2017, from <http://whatismyipaddress.com/geolocation-accuracy>
- Freni, D., Ruiz Vicente, C., Mascetti, S., Bettini, C., & Jensen, C. S. (2010). Preserving Location and Absence Privacy in Geo-social Networks. Proceedings of the 19th ACM International Conference on Information and Knowledge Management, 309–318. <https://doi.org/10.1145/1871437.1871480>
- Instagram, API Endpoints • Instagram Developer Documentation. Retrieved from: <https://www.instagram.com/developer/endpoints/> [Accessed December 6, 2016].
- Instagram, Privacy Policy, n.d. • Instagram. Retrieved from: <https://www.instagram.com/about/legal/privacy/?hl=nl#section2>. [Accessed December 6, 2016].
- Facebook, Data Policy., n.d. Retrieved from: <https://www.facebook.com/privacy/explanation>. [Accessed December 6, 2016].
- Friedland, G., & Sommer, R. (2010). Cybercasing the Joint: On the Privacy Implications of Geo-Tagging. *HotSec*, 1-6.
- Furini, M., & Tamanini, V. (2014). Location privacy and public metadata in social media platforms: attitudes, behaviors and opinions. *Multimedia Tools and Applications*, 74(21), 9795–9825. <https://doi.org/10.1007/s11042-014-2151-7>
- Gerlach, J., Widjaja, T., & Buxmann, P. (2015). Handle with care: How online social network providers' privacy policies impact users' information sharing behavior. *The Journal of Strategic Information Systems*, 24(1), 33–43. <https://doi.org/http://dx.doi.org/10.1016/j.jsis.2014.09.001>
- Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), 211-221.
- Graham, M. & Shelton, T., 2013. Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography*, 3(3), pp.255–261. <http://dhg.sagepub.com/lookup/doi/10.1177/2043820613513121>.
- asan, S., Zhan, X., & Ukkusuri, S. V. (2013, August). Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. In *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing* (p. 6). ACM.
- Hu, Y., Manikonda, L. & Kambhampati, S., 2014. What we Instagram: a first analysis of Instagram photo content and user types. Proceedings of the Eight International AAAI Conference on Weblogs and Social Media, pp.595–598.
- Jiang, Z., Heng, C. S., & Choi, B. C. F. (2013). Privacy concerns and privacy-protective behaviour in synchronous online social interactions. *Information Systems Research*, 24(3), 579–595. <https://doi.org/10.1287/isre.1120.0441>

- Jin, L., Long, X., & Joshi, J. B. D. (2012). Towards Understanding Residential Privacy by Analyzing Users' Activities in Foursquare, 25–32. <https://doi.org/10.1145/2382416.2382428>
- Kar, B., & Ghose, R. (2014). Is my information private? Geo-privacy in the World of Social Media. CEUR Workshop Proceedings, 1273(Webster), 28–31.
- Kokolakis, S. (2015). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, (July 2015), 1–29. <https://doi.org/10.1016/j.cose.2015.07.002>
- Krumm, J. (2008). A Survey of Computational Location Privacy.
- Lane, N., & Walton, N. (2008). Mobile social networking. (A. Chin & D. Zhang, Eds.). Springer. Retrieved from http://www.telecoms.com/wp-content/uploads/2009/05/buongiorno_final-fmt_nl-3110-f.pdf
- Lee, H., Park, H., & Kim, J. (2013). Why do people share their context information on social network services? a qualitative study and an experimental study on users' behavior of balancing perceived benefit and risk. *International Journal of Human Computer Studies*, 71(9), 862–877. <https://doi.org/10.1016/j.ijhcs.2013.01.005>
- Li, H., Zhu, H., Du, S., Liang, X., & Shen, X. (2016). Privacy Leakage of Location Sharing in Mobile Social Networks: Attacks and Defense. *IEEE Transactions on Dependable and Secure Computing*, 1–1. <https://doi.org/10.1109/TDSC.2016.2604383>
- Li, N., & Chen, G. (2010). Sharing location in online social networks. *IEEE network*, 24(5).
- Lindqvist, J., Cranshaw, J., Wiese, J., Hong, J., & Zimmerman, J. (2011, May). I'm the mayor of my house: examining why people use foursquare—a social-driven location sharing application. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2409–2418). ACM.
- Long, X., Jin, L., & Joshi, J. (2012, September). Exploring trajectory-driven local geographic topics in foursquare. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (pp. 927–934). ACM.
- Manikonda, L., Hu, Y. & Kambhampati, S., 2014. Analyzing User Activities, Demographics, Social Network Structure and User-Generated Content on Instagram, 1-6. Retrieved from: <http://arxiv.org/abs/1410.8099>.
- Martijn, M., & Tokmetzis, D. (2017). Je hebt wèl iets te verbergen over het levensbelang van privacy. Amsterdam: De Correspondent.
- Mascetti, S., Freni, D., Bettini, C., Wang, X. S., & Jajodia, S. (2011). Privacy in geo-social networks: Proximity notification with untrusted service providers and curious buddies. *VLDB Journal*, 20(4), 541–566. <https://doi.org/10.1007/s00778-010-0213-7>
- Nissenbaum, H. (2011). A Contextual Approach to Privacy Online. *Dædalus*, 140(4), 32–48. https://doi.org/10.1162/DAED_a_00113
- Olteanu, A. M., Huguenin, K., Humbert, M., & Hubaux, J. P. (2016). The Sharing Game: Benefits and Privacy Implications of (Co)-Location Sharing with Interdependences (No. EPFL-WORKING-218755). -.
- Pavlou, P. A. (2011). State of the information privacy literature: where are we now and where should we go? *MIS Quarterly* 35(4), 977–988.

- Piwek, L., & Joinson, A. (2016). "What do they snapchat about?" Patterns of use in time-limited instant messaging service. *Computers in Human Behavior*, 54, 358-367.
- Pötzsch, S. (2009). Privacy awareness: A means to solve the privacy paradox? 4th IFIP WG 9.2, 9.6/11.6, 11.7/FIDIS International Summer School, 2008, 298(216483), 226–236. https://doi.org/10.1007/978-3-642-03315-5_17
- Puttaswamy, K. P. N., Wang, S., Steinbauer, T., Agrawal, D., El Abbadi, A., Kruegel, C., & Zhao, B. Y. (2014). Preserving location privacy in geosocial applications. *IEEE Transactions on Mobile Computing*, 13(1), 159–173. <https://doi.org/10.1109/TMC.2012.247>
- Rahman, A. (2016) The influence of social identity when digitally sharing location. PhD thesis, University of Nottingham.
- Reid, F., & Harrigan, M. (2013). Security and Privacy in Social Networks. *Security and Privacy in Social Networks*, 197–203. <https://doi.org/10.1007/978-1-4614-4139-7>
- Rossi, L., Williams, M. J., Stich, C., & Musolesi, M. (2015). Privacy and the City: User Identification and Location Semantics in Location-Based Social Networks. *arXiv*, 387–396. <https://doi.org/10.4018/978-1-4666-1981-4.ch015>
- Roxin, A., Gaber, J., Wack, M., & Nait-Sidi-Moh, A. (2007, November). Survey of wireless geolocation techniques. In *Globecom Workshops, 2007 IEEE* (pp. 1-9). IEEE.
- Jin, K., & Thakuriah, P. (2016). The trade-off between privacy and geographic data resolution. a case of GPS trajectories combined with the social survey results. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 535-542.
- Scellato, S., Mascolo, C., Musolesi, M., and Latora, V. (2010). Distance matters: Geo-social metrics for online social networks. *Proceedings of the 3rd Workshop on Online Social Networks*, 8–8. Retrieved from <http://portal.acm.org/citation.cfm?id=1863198> \n <http://dl.acm.org/citation.cfm?id=1863198>
- Sloan, L. & Morgan, J., 2015. Who tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on twitter. *PLoS ONE*, 10(11), 1–15.
- Smith, A., O'Hara, K., & Lewis, P. (2011). Visualising the past: Annotating a life with linked open data. In *Proceedings of the 3rd International Web Science Conference* (p. 16). ACM.
- Steijn, W. M., Schouten, A. P., & Vedder, A. H. (2016). Why concern regarding privacy differs: The influence of age and (non-) participation on Facebook. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 10(1).
- Suh, J. & Aburizaiza, A. (2016). *Harvesting and Visualizing Twitter data*. Presentation, Virginia, USA. Retrieved from: http://dataservices.gmu.edu/files/Harvesting-and-Visualizing-Twitter-Data_2016.pdf
- Tang, K. P., Lin, J., Hong, J. I., Siewiorek, D. P., & Sadeh, N. (2010, September). Rethinking location sharing: exploring the implications of social-driven vs. purpose-driven location sharing. In *Proceedings of the 12th ACM international conference on Ubiquitous computing* (pp. 85-94).
- Toch, E., Cranshaw, J., Drielsma, P. H., Tsai, J. Y., Kelley, P. G., Springfield, J., .& Sadeh, N. (2010, September). Empirical models of privacy in location sharing. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, 129-138.

- Trochim, W. (2017). *Nonprobability Sampling*. [online] Socialresearchmethods.net. Retrieved from: <https://www.socialresearchmethods.net/kb/samprnon.php> [Accessed 11 January 2017].
- Tubaro, P., Casilli, A. A., & Sarabi, Y. (2014). Against the Hypothesis of the End of Privacy. (A. Bounfour, Ed.). Orsay: Springer. <https://doi.org/10.1007/978-3-319-02456-1>
- Twitter Developers. (2016). *Twitter Developers*. [online] Available at: <https://dev.twitter.com/> [Accessed 9 November. 2016].
- Twitter Help Center. n.d. FAQs about the tweet location feature. [online]. Retrieved from <https://support.twitter.com/articles/78525-faqs-about-tweet-location> [Accessed 1 December 2016]
- Twitter, Privacy Policy, n.d. Twitter. [online] Retrieved at: <https://twitter.com/privacy>. [Accessed 1 December 2016]
- Van der Veer, N., Boekee, S., Peters, O. (2017). Nationale social media onderzoek 2016. Amsterdam: Newcom Research & Consultancy.
- van Loenen, B., de Jong, J., & Zevenbergen, A. (2008). Locating mobile devices - Balancing privacy and national security.
- van Loenen, B., Kulk, S., & Ploeger, H. (2016). Data protection legislation: A very hungry caterpillar. The case of mapping data in the European Union. *Government Information Quarterly*, 33(2), 338–345. <http://doi.org/10.1016/j.giq.2016.04.002>
- Van Loenen, B., & Zevenbergen, J. (2007). Privacy (regimes) do not threaten location technology development. *Proceedings - IEEE International Conference on Mobile Data Management*, 238–242. Retrieved from: <http://doi.org/10.1109/MDM.2007.50>
- Vocht, A. (2011), *Syllabus Statistiek*. Utrecht : University Utrecht
- Wagner, D., Lopez, M., Doria, A., Pavlyshak, I., Kostakos, V., Oakley, I., & Spiliotopoulos, T. (2010). Hide and Seek: Location Sharing Practices with Social Media. *Public Policy*, 40, 55–58. <https://doi.org/10.1145/1851600.1851612>
- Young, A. L., & Quan-Haase, A. (2009). Information revelation and internet privacy concerns on social network sites: A case study of Facebook. *C&T '09*, 265–274. Retrieved from: <https://doi.org/10.1145/1556460.1556499>
- Zafeiropoulou, A. M., Millard, D. E., Webber, C., & O'Hara, K. (2013). Unpicking the privacy paradox: Can structuration theory help to explain location-based privacy decisions? *Proceedings of the 5th Annual ACM Web Science Conference*, 463–472. <https://doi.org/10.1145/2464464.2464503>

Screenshots

- Instagram (2017). Screenshot of user profile, anonymous. [Accessed 2 December 2016]
- Facebook (2017). Screenshot of user profile, anonymous. [Accessed 2 December 2016]
- Twitter (2017). Screenshot of user profile, anonymous. [Accessed 2 December 2016]

Appendices

Appendix 1 Online survey.....	71
Appendix 2 Response rate and missing values.....	78
Appendix 3 Analysis Scheme	81
Appendix 4 Likert Scale.....	85
Appendix 5 Item analysis for multiple response questions.....	92

Appendix 1 Online survey

ThesisTools

Maak en verspreid gratis je online enquête op www.thesistools.com

Beste respondent,

Graag nodigen wij u uit deel te nemen aan dit onderzoek over het gebruik van sociale media en locatie privacy. Sociale media is een verzamelbegrip voor sociale internet sites waar gebruikers de inhoud verzorgen op hun profielen of op een groepspagina. Bekende voorbeelden van sociale media zijn Facebook, Twitter, Google+, Instagram en SnapChat. Gebruikers delen hun ervaringen, verhalen en emoties met medegebruikers op het internet. Dit doen zij door foto's te publiceren, video's te lanceren of (korte) teksten te schrijven. Tegenwoordig kunnen gebruikers hun activiteiten, foto's en status updates koppelen aan locatiegegevens. Dit staat ook wel bekend als geotagging.

Ondanks de vele voordelen van sociale media brengt het ook zorgen met zich mee over de privacy van de gebruiker. Privacy blijft een gevoelig en lastig onderwerp. Op basis van dit onderzoek hopen wij een beter inzicht te krijgen in de mening van de gebruikers over sociale media en privacy.

De resultaten en uw gegevens worden niet verstrekt aan derde partijen en anoniem verwerkt volgens de gedragscode Code&Statistiek, gebaseerd op artikel 25 Wet bescherming persoonsgegevens.

Het invullen van de enquête duurt hooguit 10 minuten.

Als dank voor uw medewerking worden er twee VVV-bonnen t.w.v. 15 euro en 5 cadeaubonnen voor een gratis motorrijles van 2 uur verloot onder de respondenten. Als u geïnteresseerd bent om aan deze verloten mee te doen, kunt u uw e-mailadres en naam invullen aan het einde van de enquête.

Bedankt voor uw aandacht en medewerking!

Start

ThesisTools

Maak en verspreid gratis je online enquête op www.thesistools.com

ThesisTools

Maak en verspreid gratis je online enquête op www.thesistools.com

1. **Wat is uw geslacht?**

- Man
 Vrouw

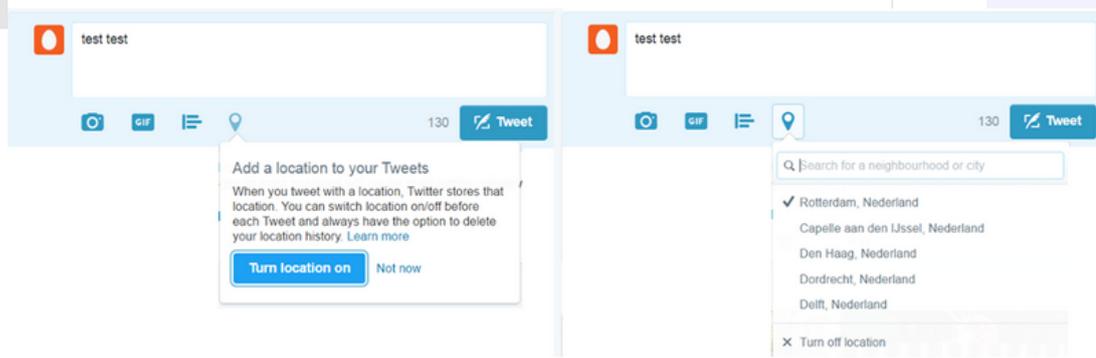
2. **Wat is uw leeftijd?**

3. **Wat is uw hoogst afgeronde opleiding?**

- Lager onderwijs
 Lager beroepsonderwijs
 VMBO
 HAVO
 VWO
 MBO
 HBO
 WO
 Anders, namelijk

Deel 2 Geotaggen op sociale media

Tegenwoordig hebben de meeste mobiele apparaten een camera- en microplatform met telefoonfunctie, waarmee mensen foto's maken en video's opnemen van hun activiteiten of van hun omgeving. Daarnaast is de sociale media toegankelijk voor gebruik op mobiele apparaten via verschillende applicaties. Sociale mediagebruikers delen hun foto's of video's op de sociale hun vrienden. Zoals eerder vermeld is het vaak ook mogelijk om de inhoud van de gebruiker te geotaggen, dit betekent het koppelen van locatie informatie aan uw sociale media berichten op uw profiel. Een paar voorbeelden hiervan zijn inchecken bij een restaurant op Facebook, een locatie toevoegen aan uw foto's op uw Instagram account of de locatie instellingen aanzetten van uw Twitteraccount. Hieronder ziet u een voorbeeld bij Twitter:



4. Hoe vaak gebruikt u de volgende sociale media?

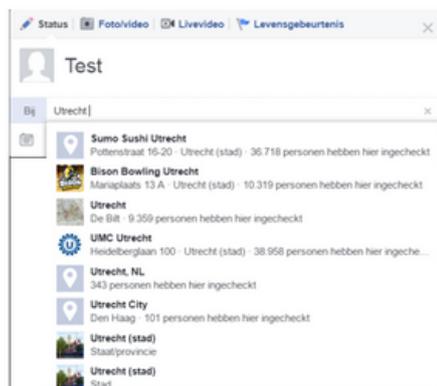
	Nooit	Zelden	Soms	Regelmatig	Bijna altijd
Facebook	<input type="radio"/>				
Twitter	<input type="radio"/>				
Instagram	<input type="radio"/>				
SnapChat	<input type="radio"/>				
Flickr	<input type="radio"/>				

5. Hoe vaak gebruikt u sociale media op de volgende apparaten?

	Nooit	Zelden	Soms	Regelmatig	Bijna altijd
Smart Phone	<input type="radio"/>				
Tablet	<input type="radio"/>				
Computer/ Laptop	<input type="radio"/>				

De gebruiker op sociale media kan kiezen tussen twee opties voor het koppelen van locatie informatie aan zijn of haar status op een sociale media platform. De eerste optie is keuze maken vanuit de lijst met suggesties (plaatje A). De tweede keuze is zelf de locatie toevoegen door een formulier in te vullen (plaatje B).

A: Lijst met suggesties van locatie



B: Plaats toevoegen met een formulier

The screenshot shows a form titled 'Plaats toevoegen' (Add location). The form contains the following fields:

- Naam (Name):
- Adres (Address):
- Plaats (Location):
- Categorieën (Categories): (Label: Wat voor type plaats is dit?)
- Website (optioneel) (Website (optional)):
- Locatie op kaart (Location on map): A map showing a street grid with a red pin. Below the map, it says 'Sleep de pin naar de locatie op de kaart' (Drag the pin to the location on the map).

At the bottom right of the form, there are two buttons: 'Opslaan' (Save) and 'Annuleren' (Cancel).

6. Van welke geotag functies maakt u gebruik op sociale media?

- Lijst met suggesties van locaties
- Plaats toevoegen met een formulier
- Beide
- Geen van beide
- Weet ik niet

7. Hoe vaak voegt u uw locatie informatie toe op uw profiel bij de volgende opties?

	Nooit	Zelden	Soms	Regelmatig	Bijna altijd
Foto	<input type="radio"/>				
Video	<input type="radio"/>				
Status update	<input type="radio"/>				
Tweet	<input type="radio"/>				

8. Bij welke locaties voegt u vooral locatie gegevens toe aan een foto op sociale media? (Meerdere antwoorden mogelijk)

- Vakantie
 Uitgaan
 Thuis
 Werk/thuis
 Anders, namelijk

9. Om welke redenen voegt u locatie gegevens toe aan de foto, tekst of video op uw profiel?

Deel 3 Houding tegenover locatie privacy en voorkeur

10. Heeft u uw privacy instellingen van uw profielen op sociale media platformen aangepast?

- Ja, altijd
 Sommige accounts wel, sommige accounts niet
 Ik heb altijd de standaardinstellingen
 Niet van toepassing

11. In hoeverre bent u tevreden mee met de opties van privacy instellingen op sociale media?

- Zeer tevreden
 Tevreden
 Neutraal
 Niet tevreden
 Zeer ontevreden
 Geen mening

12. Wie mogen welke locatie informatie van u zien op uw sociale media profiel? (Meerdere antwoorden mogelijk)

	Niemand	Familie	Vrienden	Kennissen	Collega's op werk/ studiegenoten	Werkgever
Huisadres	<input type="checkbox"/>	<input type="checkbox"/>				
Werkplek /Studieplek	<input type="checkbox"/>	<input type="checkbox"/>				
Uitgaansgelegenheid	<input type="checkbox"/>	<input type="checkbox"/>				
Ziekenhuis	<input type="checkbox"/>	<input type="checkbox"/>				
Politieke bijeenkomst	<input type="checkbox"/>	<input type="checkbox"/>				

13.

Welke partijen mogen wat u betreft uw locatie informatie gebruiken?(Meerdere antwoorden mogelijk)

	Huisadres	Werkplek of studieplek	Uitgaansgelegenheid	Ziekenhuisbezoek	Politieke bijeenkomst	Niet
Sociale media voor verbetering van eigen diensten	<input type="checkbox"/>					
Adverteerders voor gepersonaliseerde reclame op internet	<input type="checkbox"/>					
Bedrijven voor analyses en diensten	<input type="checkbox"/>					
Onderzoeksinstellingen en universiteiten voor onderzoek	<input type="checkbox"/>					
Overheidsinstanties	<input type="checkbox"/>					
Inlichtingsdiensten	<input type="checkbox"/>					

14.

Hieronder leest u beschrijving van drie personen. Welke beschrijving past het beste bij u?

Persoon A maakt zich zorgen om zijn privacy op het internet en weigert persoonlijke gegevens te delen met applicaties op zijn telefoon. Daarnaast wil deze persoon zo min mogelijk informatie delen over zijn persoonlijke leven op sociale media.

Persoon B maakt zich ook zorgen om zijn privacy, maar heeft er vrijwel geen moeite mee om een foto te delen op sociale media. Ook geeft hij toestemming aan sommige applicaties om persoonlijke gegevens op te slaan.

Persoon C vindt dat hij niets te verbergen heeft en heeft er geen probleem mee wanneer bedrijven zoals Google of Facebook zijn persoonlijke gegevens bewerken voor eigen doeleinden.

- Persoon A
 Persoon B
 Persoon C
 Geen van allen
 Weet ik niet

15. In hoeverre bent u het eens met de volgende stellingen?

	Helemaal mee oneens	Oneens	Neutraal	Eens	Helemaal mee eens	Weet niet
Ik maak mij zorgen om inbreuk op mijn privacy door hackers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ik maak mij zorgen om reputatieschade binnen mijn sociale kringen en op de werkvloer door inbreuk op mijn locatie privacy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ik maak mij zorgen om identiteitsfraude door inbreuk op mijn locatie privacy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ik maak mij zorgen om het verkopen van mijn locatie informatie aan adverteerders	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ik maak mij zorgen om het verkopen van mijn locatie informatie aan bedrijven	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ik maak mij zorgen om het gebruik van mijn locatie informatie door overheidsinstanties	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ik maak mij zorgen om het gebruik van mijn locatie informatie door universiteiten en onderzoeksinstituten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

16. Als dank voor uw medewerking verloten wij 2 vvv giftcards ter waarde van 15 euro en 5 motorrijles cadeaubonnen onder de respondenten.

Mocht u geïnteresseerd zijn in de loten dan kunt u hier uw naam en mailadres invullen.

17. In welke loten bent u geïnteresseerd?

- VVV giftcard ter waarde van 15 euro
 Cadeaubon motorrijles van 2 uur
 Beide
 Geen interesse

Appendix 2 Response rate and missing values

Variables	N	Missing	
		Count	Percent
age	179	5	2,7
gender	181	3	1,6
education	181	3	1,6
Social media			
Facebook	157	27	14,7
Twitter	157	27	14,7
Instagram	155	29	15,8
SnapChat	157	27	14,7
Flickr	155	29	15,8
Mobile device			
SmartPhone	155	29	15,8
Tablet	155	29	15,8
ComputerLaptop	156	28	15,2
Geotag feature	157	27	14,7
Content			
Foto	158	26	14,1
Video	157	27	14,7
Statusupdate	157	27	14,7
Tweet	156	28	15,2
Location			
Vacation	161	23	12,5
Trips	161	23	12,5
Home	161	23	12,5
Workplace	161	23	12,5
Other	161	23	12,5
instellingaanpassen	132	52	28,26087
tevredenheidinstelling	129	55	29,8913
Audience			
HuisadresNiemand	139	45	24,5
HuisadresFamilie	139	45	24,5
HuisadresVrienden	139	45	24,5
HuisadresKennissen	139	45	24,5
HuisadresCollega'sopwerkstudiegenoten	139	45	24,5
HuisadresWerkgever	139	45	24,5
WerkplekStudieplekNiemand	139	45	24,5

WerkplekStudieplekFamilie	139	45	24,5
WerkplekStudieplekVrienden	139	45	24,5
WerkplekStudieplekKennissen	139	45	24,5
WerkplekStudieplekCollega'sopwerkstudiegenoten	139	45	24,5
WerkplekStudieplekWerkgever	139	45	24,5
UitgaansgelegenheidNiemand	139	45	24,5
UitgaansgelegenheidFamilie	139	45	24,5
UitgaansgelegenheidVrienden	139	45	24,5
UitgaansgelegenheidKennissen	139	45	24,5
UitgaansgelegenheidCollega'sopwerkstudiegenoten	139	45	24,5
UitgaansgelegenheidWerkgever	139	45	24,5
ZiekenhuisNiemand	139	45	24,5
ZiekenhuisFamilie	139	45	24,5
ZiekenhuisVrienden	139	45	24,5
ZiekenhuisKennissen	139	45	24,5
ZiekenhuisCollega'sopwerkstudiegenoten	139	45	24,5
ZiekenhuisWerkgever	139	45	24,5
PolitiekbijeenkomstNiemand	139	45	24,5
PolitiekbijeenkomstFamilie	139	45	24,5
PolitiekbijeenkomstVrienden	139	45	24,5
PolitiekbijeenkomstKennissen	139	45	24,5
PolitiekbijeenkomstCollega'sopwerkstudiegenoten	139	45	24,5
PolitiekbijeenkomstWerkgever	139	45	24,5
Reuse			
SocialemediavoorverbeteringvaneigendienstenHuisadres	139	45	24,5
SocialemediavoorverbeteringvaneigendienstenWerkplekofstudieplek	139	45	24,5
SocialemediavoorverbeteringvaneigendienstenUitgaansgelegenheid	139	45	24,5
SocialemediavoorverbeteringvaneigendienstenZiekenhuisbezoek	139	45	24,5
SocialemediavoorverbeteringvaneigendienstenPolitiekbijeenkomst	139	45	24,5
SocialemediavoorverbeteringvaneigendienstenNiet	139	45	24,5
AdverteerdersvoorgepersonaliseerdereclameopinternetHuisadres	139	45	24,5
AdverteerdersvoorgepersonaliseerdereclameopinternetWerkplekofstu	139	45	24,5
AdverteerdersvoorgepersonaliseerdereclameopinternetUitgaansgeleg	139	45	24,5
AdverteerdersvoorgepersonaliseerdereclameopinternetZiekenhuisbez	139	45	24,5
AdverteerdersvoorgepersonaliseerdereclameopinternetPolitiekbije	139	45	24,5
AdverteerdersvoorgepersonaliseerdereclameopinternetNiet	139	45	24,5
BedrijvenvooranalysesendienstenHuisadres	139	45	24,5
BedrijvenvooranalysesendienstenWerkplekofstudieplek	139	45	24,5

BedrijvenvooranalysesendienstenUitgaansgelegenheid	139	45	24,5
BedrijvenvooranalysesendienstenZiekenhuisbezoek	139	45	24,5
BedrijvenvooranalysesendienstenPolitiekebijeenkomst	139	45	24,5
BedrijvenvooranalysesendienstenNiet	139	45	24,5
Onderzoeksinstellingen en universiteiten voor onderzoek Huisadres	139	45	24,5
Onderzoeksinstellingen en universiteiten voor onderzoek Werkplek of stu	139	45	24,5
Onderzoeksinstellingen en universiteiten voor onderzoek Uitgaansgeleg	139	45	24,5
Onderzoeksinstellingen en universiteiten voor onderzoek Ziekenhuisbez	139	45	24,5
Onderzoeksinstellingen en universiteiten voor onderzoek Politieke bije	139	45	24,5
Onderzoeksinstellingen en universiteiten voor onderzoek Niet	139	45	24,5
Overheidsinstanties Huisadres	139	45	24,5
Overheidsinstanties Werkplek of studieplek	139	45	24,5
Overheidsinstanties Uitgaansgelegenheid	139	45	24,5
Overheidsinstanties Ziekenhuisbezoek	139	45	24,5
Overheidsinstanties Politieke bijeenkomst	139	45	24,5
Overheidsinstanties Niet	139	45	24,5
Inlichtingsdiensten Huisadres	139	45	24,5
Inlichtingsdiensten Werkplek of studieplek	139	45	24,5
Inlichtingsdiensten Uitgaansgelegenheid	139	45	24,5
Inlichtingsdiensten Ziekenhuisbezoek	139	45	24,5
Inlichtingsdiensten Politieke bijeenkomst	139	45	24,5
Inlichtingsdiensten Niet	139	45	24,5
Attitude privacy interest position	135	49	26,6
Concerns			
Concerns hackers	139	45	24,5
Concerns problems	138	46	25,0
Concerns identity	138	46	25,0
Concerns adverteerders	135	49	26,6
Concern bedrijf	133	51	27,7
Concern Overheid	137	47	25,5
Concern universiteit	137	47	25,5

Appendix 3 Analysis Scheme

The analysis scheme gives an overview of the used descriptive statistic methods. The survey questions create many categoric variables that are useful for examining the associations between factors such as type of places and geotagging of content. With cross tables and correlations coefficients it's possible to determine whether if there is a significant association between two or more variables. The measure scale of a variable determines which correlation coefficient is suitable to measure the significant and strength of the association.

Several cross tables are made to confirm to examine the values of variables and their association between. Chi-square tests are executed to examine whether there is a significant association between the factors. To determine if there is a weak or strong association Cramer's V, Kendall's Tau and Spearman's Rho as correlation coefficient are also applied for the cross tabulations. The cross tabulations are available in the Appendix.

Pearson's Chi-Square test

Pearson's Chi-Square test also known as simply Chi-square test is a statistical hypothesis test wherein the sampling distribution is compared to the chi-squared distribution. This means the observed cell frequencies are compared to the expected cell frequencies of the cross table. The null hypothesis assumes there is no significant difference between the expected frequencies and the observed frequencies in one or more categories. The differences between observed and expected cell frequencies are by chance and there is no statistical association (Vocht, 2011). However, if there is a significant difference between the expected and observed cell frequencies, there is a statistical association between one or more categories. If the asymptotic significance level is less than 5% ($\alpha = 0.05$), the null hypothesis, which is not statistically related, is rejected.

Correlation coefficients: Cramer's V and Kendall rank and Spearman's Rho correlation coefficient

Correlation coefficients indicates if there is an association between two or more variables. If particularly values are frequently common between two variables, then there is an association. The association is determined by its strength and direction. For the interpretation of the association the rule of Cohen will be used as guideline:

0-0,10	Very weak
0,11-0,30	Weak
0,31-0,50	Moderately strong
0,51-0,80	Strong
0,80-0,99	Very strong
1	Perfect association

Cramer's V is one of the most used correlation coefficient, for the relationship between two nominal variables. It's also used to indicate the relation between nominal and weak ordinal variables. Cramer's V is an association measurement based on Chi-Square test, which measure only the strength of the association, and not the causality between an independent and dependent variable.

Kendall rank correlation coefficient also known as Kendall's tau coefficient is a suitable correlation measure for weak ordinal variables and measure rank correlation. Which means the similarity of the orderings of values when its ranked-on quantities. The coefficient will be positively high when the observed values have a similar rank between the compared variables. However, the coefficient will be low when the observed values have a dissimilar rank. Just as the Cramer's V, Kendall tau correlation doesn't indicate a causality between two measured variables.

Spearman's Rho is a correlation coefficient between rang order numbers of ordinal, interval and ratio measures. In contrast to Cramer's V and Kendall's Tau, Spearman's Rho is not based on a cross table, but based on the differences between the rang order numbers. Rho will be zero when there is no association between the variables, whereas the Rho between -1 and 1 refers to an association. It's also possible to calculate the common variance of the variables.

The null hypothesis assumes there is no significant difference between rang order numbers of two variables. The differences between rang order numbers are by chance and there is no statistical association between the ordinal variables (Vocht, 2011). If the asymptotic significance level is less than 5% ($\alpha = 0.05$), the null hypothesis, which is not statistically related, is rejected.

Privacy paradox

Privacy paradox is defined as the dichotomy between the participant's usage of geotag feature on social media and their concerns and attitudes towards location privacy. Attitudes and concerns are not the same, even though both are similar and influence each other. For this reason, both factors will be analysed separately in relation to the willingness to geotag content on social media.

First, the geotagging behaviour will be defined, whereas the attitude will be based on the question 14 about the attitude descriptions. The concerns component will be also redefined to compare with the geotagging behaviour of participants.

The geotagging behaviour is expressed in two ways: use of geotag features and the willingness to geotag. The use of geotag feature is a variable based on question 6. The options "list with suggestions", "Add places manually" and "Both" are compromised as one category (0) because it means the participant add location to their content on their profiles. Whereas the option "None" is categorized as category (0), which means the participant doesn't geotag on social media.

The willingness to geotag is based on question 7 about how many times participants geotag their photos, videos, status updates and tweets online. Since question 7 is Likert-scale, it is possible to work with numerical variables. The Likert-scale options go as follows: Never (1), Rarely (2), Sometimes (3), Regularly (4) and Continuously (5). A new variable was introduced as Willingness to geotag, which was the score of each participant's answer in question 7 (Video, Photo, Status update and Tweet). The score was calculated based on the answers given by the participant in each item divided by the number of answer. The association between the willingness of geotag and the user's characteristics gender, age and education level are calculated with Cramer's V and Kendal's Tau (Table 3-1 till 3-4).

Attitude

The Cronbach's Alpha test is conducted to examine intern consistency of the multiple response question 12 and 13 (Table 3-5 and Table 3-6). The results of Cronbach's Test are presented in Appendix 4. The association between the privacy interest position and participant's willingness to geotag is calculated with Cramer's V. The association between the concerns of the participants and their geotagging behaviour is calculated with Spearman's Rho (Table 3-7).

Concerns

Question 15 is a Likert-scale with 7 items. The 7 items measure the participants concerns related to monetization of personal data and re-use of location data. As mentioned in chapter 3, the participants were asked to indicate to which extent they agree with the statements with these options: "Strongly disagree", "Disagree", "Agree", "Neutral", "Agree" and "Strongly agree". The options are numbered from 1 till 6. Number 1 refers to "Strongly disagree", whereas number 6 refers to "Strongly agree" and other numbers refers to the options between. The mean Likert score of each participant is calculated with the mean for two scales threats (hackers, social issues, identity theft) and concerns related to re-

use of location information (companies, government, research and intelligence service). The association between participant's concern and their willingness to geotag content is calculated with Spearman's Rho.

Table 3-1 Cross table: Use of geotag feature & Location types

Variable	Geotag feature
Survey question	6
Measure scale	Nominal
Variable	Content
Survey question	8
Measure scale	Nominal
Analysis	Chi-Square, Cramer's V

Table 3-2 Cross table: gender willingness to geotag

Variable	Geotag feature
Survey question	6
Measure scale	Nominal
Variable	Gender
Survey question	1
Measure scale	Nominal
Analysis	Chi-Square, Cramer's V

Table 3-3 Cross table: age classes and willingness to geotag

Variable	Geotag feature
Survey question	6
Measure scale	Nominal
Variable	Age classes
Survey question	2
Measure scale	Ordinal
Analysis	Chi-Square, Cramer's V

Table 3-4 Cross table: education classes and willingness to geotag

Variable	Geotag feature
Survey question	6
Measure scale	Nominal
Variable	Age classes
Survey question	2
Measure scale	Ordinal
Analysis	Chi-Square, Cramer's V

Table 3-5 Audience on social media

Variable	Audience
Survey question	12
Measure scale	Nominal
Frequency	Table, Bar
Analysis	Cronbach's Alpha

Table 3-6 Reuse of location information by third parties

Variable	Audience
Survey question	13
Measure scale	Nominal
Frequency	Table Bar
Analysis	Cronbach's Alpha

Table 3-7 Cross Table: Willingness to geotag & Attitude position

Variable	Geotagging content
Survey question	7
Measure scale	Ordinal
Variable	Attitude privacy interest position
Survey question	15
Measure scale	Nominal
Analysis	Cramer's V

Table 3-8 Cross Table: Willingness to geotag & Concerns (Mean)

Variable	Geotagging content
Survey question	7
Measure scale	Ordinal
Variable	Concerns
Survey question	15
Measure scale	Ordinal
Analysis	Chi-Square, Kendall's Tau

Appendix 4 Likert Scale

Likert-scale is a psychometric method to scale responses in surveys with ratings to measure a latent variable. The latent variable is the underlying phenomenon which may be an opinion on a matter or a social behaviour which might be hard to with the help of diverse manifest variables also known as items it is possible to measure the latent variable. The latent variable is operationalized into manifest variables that capture the underlying phenomenon. To indicate if the items of the Likert-Scale measure the underlying latent variable two tests are applied, factor- and reliability analysis.

Factor analysis

To indicate if the Likert-scales measures one latent variable also known as *component* by SPSS, factor analysis is applied for the questions 4,5, 6 about social media and geotagging and question 15 about concerns related to location privacy. The correlation between a manifest variable and a component is defined as *factor*. When the factor has value of zero this means there is no association between the manifest and latent variable. However, when the factor is bigger than 0,45 there is an association between a manifest and the component or also known as the latent variable.

The factor can also be interpreted as a correlation coefficient and a standardized regression coefficient. The value of factor is between -1 and +1 and points out a direction. A value of +1 means perfect associations between the latent and manifest variable. By means of factors the proportion of explained variance can be determined. The variance between the manifest and latent variable can be explained by each separate manifest variable, but also as a whole.

Cronbach's Alpha

Mistakes may have been made during the construction of the Likert scale or the respondents have interpreted the questions differently than the researcher intended. These errors affect the reliability of the Likert scale and the validity of the measurements. The reliability of a measurement is the extent to which that measurement is free from accidental errors. Using Cronbach's Alpha, the number of internal consistency is indicated by several numbers. In addition to reliability, Cronbach's Alpha indicates how much the Likert scale is internally consistent. In other words, Cronbach's Alpha shows how closely related a set of items are as a group. When Cronbach's Alpha value is 0.60 or higher the Likert scale is considered internally consistent.

Constructing scales

If the tests results show that the Likert-scales are intern consistent and trustworthy Likert-scales will be reconstructed. The Likert-scales are reconstructed by calculating the mean of participant's answers. The scale has a new measure scale: interval or ordinal rang order numbers.

Use of social media

The use of social media has been measured by asking the respondents to what extent they use the following social media: Facebook, Instagram, Snapchat, Twitter and Flickr. The factor analysis with Varimax rotation showed that two components can be distinguished (Eigenvalue > 1), namely " use of social media " and " no use of social media " (Table 4-1). Together, these factors explain 55.9% of the variance (Table 4-2). Only social media usage scale is reliable ($\alpha = 0.627$) and internally consistent while scale " no use of social media " is not internally consistent ($\alpha = 0,101$) (Table 4-4 and 4-6). Participants do not use many social media (M = 2.91, SD = 1.05) and very little of Twitter and Flickr (M = 1.4 and SD = 0.70), both measured on a scale of 1 to and with 5.

Factor analysis of Likert-scale "Use of social media"

Table 4-1 Social media: Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings	
	Total	% of Variance	Cumulative %	Total	% of Variance
1	1,786	35,715	35,715	1,786	35,715
2	1,057	21,145	56,860	1,057	21,145
3	,908	18,153	75,014		
4	,820	16,395	91,408		
5	,430	8,592	100,000		

Table 4-2 Social media: Total Variance Explained

Component	Extraction Sums of Squared Loadings	Rotation Sums of Squared Loadings		
	Cumulative %	Total	% of Variance	Cumulative %
1	35,715	1,739	34,784	34,784
2	56,860	1,104	22,076	56,860

Table 4-3 Social media: Rotated Component Matrix^a

	Component	
	1	2
Facebook	,558	,131
Twitter	,203	,645
Instagram	,850	,012
SnapChat	,811	,030
Flickr	-,073	,818

Reliability Analysis of Likert-scale "Use of social media"

Table 4-4 Social media Reliability Statistics

Cronbach's Alpha	N of Items
,627	3

Table 4-5 Social Media Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Facebook	4,63	7,189	,279	,711
Instagram	6,36	4,087	,552	,344
SnapChat	6,49	4,330	,519	,401

Table 4-6 No use Social Media Reliability Statistics

Cronbach's Alpha	N of Items
,101	2

Table 4-7 No use Social media: Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Twitter	1,06	,126	,096	.
Flickr	1,68	1,395	,096	.

Use of mobile device for social media

The use of mobile device for social media has been measured by asking the respondents to what extent they use the following devices: Smartphone, tablet and computer or laptop. The factor analysis with Varimax rotation showed that two components could be distinguished (Eigenvalue > 1), namely " use of mobile device " and " no use of mobile device " (Table 4-8). Together, these factors explain 70.7% of the variance (Table 4-9). Both scales are not reliable because the Cronbach Alpha has very low value ($\alpha = 0.120$) (Table 4-11). In general participants use smartphones and laptops ($M = 3.95$ $SD = 0.79$), whereas tablets are never used or sometimes ($M = 2.7$ $SD = 0.86$).

Factor analysis of Likert-scale "Use of mobile device for social media"

Table 4-8 Mobile Device: Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings	
	Total	% of Variance	Cumulative %	Total	% of Variance
1	1,076	35,865	35,865	1,076	35,865
2	1,044	34,800	70,665	1,044	34,800
3	,880	29,335	100,000		

Table 4-9 Mobile Device: Total Variance Explained

Component	Extraction Sums of Squared Loadings	Rotation Sums of Squared Loadings		
	Cumulative %	Total	% of Variance	Cumulative %
1	35,865	1,060	35,349	35,349
2	70,665	1,059	35,316	70,665

Table 4-10 Mobile Device: Rotated Component Matrix

	Component	
	1	2
Smart Phone	,832	,193
Tablet	,183	,835
Computer/ Laptop	,578	-,569

Reliability Analysis of Likert-scale "Use of mobile device for social media"

Table 4-11 Reliability Statistics

Cronbach's Alpha	N of Items
,120	2

Table 4-12 Use of mobile device Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Smart Phone	3,50	1,331	,064	.
Computer/ Laptop	4,37	1,012	,064	.

Table 4-13 No use of mobile device Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Tablet	4,37	1,016	,065	.
Smart Phone	1,78	1,367	,065	.

Geotagging content

The participant's willingness to geotag content is measured by asking the participants to which extent they geotag the following content: photo, video, tweets and status update. The latent variable "willingness to geotag" is divided into 4 manifest variables related to the content. The answer scale is "never", "rarely", "sometimes", "regularly" and "continuously". The factor analysis with varimax rotation shows the Likert-scale measures one component, namely "willingness to geotag" (Eigenvalue>1) (Table 4-14). The component "geotagging content" explains 55,4% of the variance (Table 4-15).

For the reliability analysis 155 (84,2%) cases are considered as valid, whereas 29 (15,8%) cases are list wise excluded from the reliability analysis. Cronbach's Alpha value is 0.729 which means the items of the multiple responses is reliable and intern consistent (Table 4-17). Considering column "Cronbach's Alpha if item deleted" one of the items could be excluded from the analysis to strengthen the intern consistency of the Likert scale. In this case, item "Tweet" must be excluded from the analysis for a stronger intern consistency (Table 4-18). Participants rarely geotag their content on social media (M=1,98, SD=0,9).

Factor analysis of Likert-scale 'Geotagging content'

Table 4-14 Content: Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings	
	Total	% of Variance	Cumulative %	Total	% of Variance
1	2,218	55,444	55,444	2,218	55,444
2	,899	22,480	77,924		
3	,477	11,915	89,839		
4	,406	10,161	100,000		

Table 4-15 Content: Total Variance Explained

Component	Extraction Sums of Squared Loadings
	Cumulative %
1	55,444

Table 4-16 Content: Component Matrixa

	Component
	1
Foto	,827
Video	,833
Status update	,816
Tweet	,418

Reliability Analysis of Likert-scale "Geotagging content"

Table 4-17 Geotagging Content: Reliability Statistics

Cronbach's Alpha	N of Items
,729	4

Table 4-18 Geotagging content: Item-Total Statistics

Content	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Foto	4,57	3,584	,626	,602
Video	5,37	4,171	,640	,595
Status update	4,95	3,810	,610	,611
Tweet	5,79	6,506	,247	,786

Concerns regarding location privacy:

The latent variable "location privacy concern" is measured by asking the participants to which extent they agree with the statements about location privacy. The latent variable is divided into 7 manifest variables also known as items. The items are based on two main categories (1) privacy threats and (2) third parties who may have access to user's location information, mined from social media. The privacy threats are identity fraud, invasion of privacy by hackers and problems within social relations due sharing of personal information on web. The third parties are defined by advertisers and companies who sell location information and government and intelligence services using location information for their own services.

The findings of factor analysis show there are two components (eigenvalue>1), "privacy threats" and "re-use of location information by third party" (Table 4-19). Both components explain 65% of the variance (Table 4-20). To check if all the items measure the latent variable "location privacy concern" Cronbach's Alpha test is conducted. The before mentioned 7 items are manifestations of concerns regarding location privacy¹⁵. They were based on the trust in diverse actors such as the advertisers, companies and research institutions, but also on risk calculation of the participants on privacy threats such as identity fraud or issues in social network. These questions are the basis for the reliability analysis and enabled the exploration of the emergent factors associated with concerns. The reliability analysis is executed to confirm the internal consistency of the items. Factor analysis shows that Likert-Scales has two components, which means the reliability analysis concerns two components. The original Likert-Scale is divided into two scales.

133 cases are considered as valid, while 51 cases are list wise excluded from the reliability test. Cronbach's Alpha value for Threats is 0,700 which means the internal consistency is relatively high (Table 4-22). Considering column "Cronbach's Alpha if item deleted" one of the items could be

¹⁵ The Likert-scale results are transformed to rang numbers to determine the correlation between the use of geotag feature and participant's concerns. The options of the liker-scales are numbered from 1 till 5: 1= Strongly disagree, 2= Disagree, 3=Neutral, 4=Agree and 5= Strongly agree.

excluded from the analysis to strengthen the intern consistency of the Likert scale and the output of the cross tables. However, the values of the items are less or equal to the Cronbach's Alpha (Table and all items will be included into the cross tables for paradox (Section 9.3).

Cronbach's Alpha value for re-use is 0,835 which means the internal consistency is relatively high (Table 4-24). Considering column " Cronbach's Alpha if item deleted" one of the items could be excluded from the analysis to strengthen the intern consistency of the Likert scale and the output of the cross tables. However, the values of the items are less or equal to the Cronbach's Alpha (Table 4-25) and all items will be included into the cross tables for paradox (Section 9.3). Both components are reliable and intern consistent.

Participants are quite neutral about the privacy threats on location privacy ($M=3,12$, $SD=0,94$), but also towards to re-use of their location information by third parties such as companies and research institutions ($M=3$, $SD=0,98$), both components are measured on scale 1 till 5.

Factor analysis of Likert-scale 'Concerns regarding location privacy

Table 4-19 Concerns: Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings	
	Total	% of Variance	Cumulative %	Total	% of Variance
1	3,590	51,279	51,279	3,590	51,279
2	1,010	14,429	65,708	1,010	14,429
3	,851	12,151	77,859		
4	,687	9,814	87,673		
5	,414	5,909	93,582		
6	,290	4,139	97,721		
7	,160	2,279	100,000		

Table 4-20 Concerns: Total Variance Explained

Component	Extraction Sums of Squared Loadings	Rotation Sums of Squared Loadings		
		Cumulative %	Total	% of Variance
1	51,279	2,504	35,766	35,766
2	65,708	2,096	29,942	65,708

Table 4-21 Concerns Threats: Rotated Component Matrix

	Component	
	1	2
Ik maak mij zogen om inbreuk op mijn,262 privacy door hackers		,791

Ik maak mij zorgen om reputatieschade binnen mijn sociale kringen en op de werkvloer door inbreuk op mijn locatie privacy	,164	,636
Ik maak mij zorgen om identiteitsfraude door inbreuk op mijn locatie privacy	,184	,828
Ik maak mij zorgen om het verkopen van mijn locatie informatie aan adverteerders	,633	,483
Ik maak mij zorgen om het verkopen van mijn locatie informatie aan bedrijven	,790	,306
Ik maak mij zorgen om het gebruik van mijn locatie informatie door overheidsinstanties	,838	,186
Ik maak mij zorgen om het gebruik van mijn locatie informatie door universiteiten en onderzoeksinstellingen	,805	,132

Reliability Analysis of Likert-scale "Concerns regarding location privacy"

Table 4-22 Concerns Threats: Reliability Statistics

Cronbach's Alpha	N of Items
,700	3

Table 4-23 Concerns Threats: Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Ik maak mij zorgen om inbreuk op mijn privacy door hackers	6,03	4,117	,601	,512
Ik maak mij zorgen om reputatieschade binnen mijn sociale kringen en op de werkvloer door inbreuk op mijn locatie privacy	6,53	4,266	,404	,757
Ik maak mij zorgen om identiteitsfraude door inbreuk op mijn locatie privacy	6,12	3,913	,563	,549

Table 4-24 Concerns re-use Reliability Statistics

Cronbach's Alpha	N of Items
,835	4

Table 4-25 Concerns re-use: Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Ik maak mij zorgen om het verkopen van mijn locatie informatie aan adverteerders	8,66	9,112	,656	,796
Ik maak mij zorgen om het verkopen van mijn locatie informatie aan bedrijven	8,79	8,091	,761	,746
Ik maak mij zorgen om het gebruik van mijn locatie informatie door overheidsinstanties	9,42	9,463	,668	,791
Ik maak mij zorgen om het gebruik van mijn locatie informatie door universiteiten en onderzoeksinstellingen	9,80	10,012	,585	,825

Appendix 5 Item analysis for multiple response questions

To measure the internal consistency of multiple response questions the Cronbach's Alpha test is conducted. The test is a measure of internal consistency. In other words, Cronbach's Alpha shows how closely related a set of items are as a group. The items of a Likert scale are the manifested variables that refers to one concept, latent variable. The items of the Likert-scale have relatively low intern consistency when the value of Cronbach's Alpha is less than 0,70.

Attitude: location information audience on social media

Question 12 examines the attitude of participants towards fellow users who have access to their location information on social media. As mentioned before, privacy is a personal concept and highly depends on the context of place and the type of relationship between the sender, owner of the location information, and the receiver (the audience on social media). To examine who is permitted to view location information and its type of place the latent variable "audience" is divided into manifest variables "family", "friends", "acquaintance", "colleague" and "employer". "Nobody" is also added as item to give the participants an option if they wish that nobody can see their location information or at least to keep it more less private. The types of place are: "Home", "Trip", "Workplace/Study place", "Hospital" and "Political event".

For the reliability analysis with Cronbach's Alpha 139 (75,5%) cases are considered valid, whereas 45(24,5%) cases are list wise excluded. The Cronbach's Alpha is 0,835 which mean the items for the re-use question shows also a high internal consistency (Table 5-1). The items related to the actors measure the underlying concept and attitudes towards the re-use of personal location information by third parties such as advertisers and government. Considering column " Cronbach's Alpha if item deleted" one of the items could be excluded from the analysis to strengthen the intern consistency of the Likert scale. However, the values of the items are less or equal to the original Cronbach's Alpha (0,835) and all items will be included into the analysis (Table 5-2).

Table 5-1 Audience: Reliability Statistics

Cronbach's Alpha	N of Items
,835	30

Table 5-2 Audience: Item-Total Statistics

Place and audience	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Huisadres Niemand	8,83	30,303	-,503	,859
Huisadres Familie	9,19	25,559	,458	,826
Huisadres Vrienden	9,22	25,348	,530	,823
Huisadres Kennissen	9,38	25,904	,616	,824
Huisadres Collega's op werk/ studiegenoten	9,36	25,913	,558	,825
Huisadres Werkgever	9,40	26,487	,456	,828
Werkplek /Studieplek Niemand	9,21	30,021	-,476	,857
Werkplek /Studieplek Familie	8,94	24,931	,543	,822
Werkplek /Studieplek Vrienden	8,81	25,602	,432	,827
Werkplek /Studieplek Kennissen	9,05	24,860	,561	,821
Werkplek /Studieplek Collega's op werk/ studiegenoten	9,00	24,594	,612	,819
Werkplek /Studieplek Werkgever	9,14	25,066	,546	,822
Uitgaansgelegenheid Niemand	9,17	30,390	-,531	,859
Uitgaansgelegenheid Familie	9,04	24,948	,540	,822
Uitgaansgelegenheid Vrienden	8,83	25,342	,480	,825
Uitgaansgelegenheid Kennissen	9,17	24,675	,646	,819
Uitgaansgelegenheid Collega's op werk/ studiegenoten	9,22	24,852	,642	,819
Uitgaansgelegenheid Werkgever	9,35	25,998	,517	,825

Ziekenhuis Niemand	8,78	30,044	-,471	,857
Ziekenhuis Familie	9,24	25,664	,472	,826
Ziekenhuis Vrienden	9,29	25,525	,550	,823
Ziekenhuis Kennissen	9,40	25,763	,715	,822
Ziekenhuis Collega's op werk/ studiegenoten	9,38	25,716	,680	,822
Ziekenhuis Werkgever	9,41	26,200	,611	,825
Politieke bijeenkomst Niemand	8,88	30,943	-,602	,863
Politieke bijeenkomst Familie	9,18	24,540	,687	,817
Politieke bijeenkomst Vrienden	9,14	25,046	,547	,822
Politieke bijeenkomst Kennissen	9,31	25,085	,700	,819
Politieke bijeenkomst Collega's op werk/ studiegenoten	9,29	25,192	,637	,821
Politieke bijeenkomst Werkgever	9,38	25,774	,660	,823

Attitude: re-use of location information by third parties

Question 13 examines the attitude of participants towards third parties such as advertisers, government and companies who have access to their location information. To examine which actor is permitted to re-use location information and its type of place the latent variable "re-use of location information" is divided into manifest variables "social media", "companies", "advertisers", "intelligence services", "research and universities" and "government". "No place" is also added as item to give the participants an option if they wish that no actor should be permitted to use their location information or at least to keep it more less private. The types of place are: "Home", "Trip", "Workplace/Study place", "Hospital" and "Political event".

For the reliability analysis 139 (75,5%) cases are considered as valid, whereas 45 (24,5%) cases are list wise excluded from the reliability analysis. Cronbach's Alpha value is 0.825 which means the items of the multiple responses is highly reliable and intern consistent (Table 5-3). Considering column "Cronbach's Alpha if item deleted" one of the items could be excluded from the analysis to strengthen the intern consistency of the Likert scale. However, the values of the items are less or equal to the original Cronbach's Alpha (0,825). All items will be included into the analysis (Table 5-4).

Table 5-3 Reliability statistics

Cronbach's Alpha	N of Items
,825	36

Table 5-4 Re-use of location information by third parties: Item-Total Statistics

Place type and actor	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Sociale media voor verbetering van eigen diensten Huisadres	9,62	28,281	,252	,823
Sociale media voor verbetering van eigen diensten Werkplek of studieplek	9,50	26,672	,541	,814
Sociale media voor verbetering van eigen diensten Uitgaansgelegenheid	9,48	26,701	,520	,814
Sociale media voor verbetering van eigen diensten Ziekenhuisbezoek	9,66	27,878	,528	,818
Sociale media voor verbetering van eigen diensten Politieke bijeenkomst	9,63	27,120	,664	,814
Sociale media voor verbetering van eigen diensten Niet	9,01	31,753	-,542	,850
Adverteerders voor gepersonaliseerde reclame op internet Huisadres	9,68	28,525	,295	,823
Adverteerders voor gepersonaliseerde reclame op internet Werkplek of studieplek	9,60	27,328	,529	,816
Adverteerders voor gepersonaliseerde reclame op internet Uitgaansgelegenheid	9,56	27,451	,416	,818
Adverteerders voor gepersonaliseerde reclame op internet Ziekenhuisbezoek	9,68	28,218	,506	,820
Adverteerders voor gepersonaliseerde reclame op internet Politieke bijeenkomst	9,63	27,640	,504	,818
Adverteerders voor gepersonaliseerde reclame op internet Niet	8,90	31,280	-,522	,845

Bedrijven voor analyses en diensten	9,64	27,870	,442	,819
Huisadres				
Bedrijven voor analyses en diensten	9,47	26,425	,572	,812
Werkplek of studieplek				
Bedrijven voor analyses en diensten	9,50	26,368	,624	,811
Uitgaansgelegenheid				
Bedrijven voor analyses en diensten	9,65	27,578	,582	,817
Ziekenhuisbezoek				
Bedrijven voor analyses en diensten	9,62	27,209	,608	,815
Politieke bijeenkomst				
Bedrijven voor analyses en diensten	9,01	32,246	-,632	,852
Niet				
Onderzoeksinstellingen	en 9,47	26,628	,531	,814
universiteiten voor onderzoek				
Huisadres				
Onderzoeksinstellingen	en 9,21	26,036	,559	,812
universiteiten voor onderzoek				
Werkplek of studieplek				
Onderzoeksinstellingen	en 9,36	25,841	,631	,809
universiteiten voor onderzoek				
Uitgaansgelegenheid				
Onderzoeksinstellingen	en 9,48	26,512	,565	,813
universiteiten voor onderzoek				
Ziekenhuisbezoek				
Onderzoeksinstellingen	en 9,46	25,757	,724	,807
universiteiten voor onderzoek				
Politieke bijeenkomst				
Onderzoeksinstellingen	en 9,27	32,487	-,631	,855
universiteiten voor onderzoek Niet				
Overheidsinstanties Huisadres	9,42	26,404	,539	,813
Overheidsinstanties Werkplek of studieplek	9,36	25,783	,644	,809
Overheidsinstanties	9,50	26,237	,657	,810
Uitgaansgelegenheid				

Overheidsinstanties		9,50	26,179	,672	,809
Ziekenhuisbezoek					
Overheidsinstanties	Politieke	9,53	26,469	,627	,811
bijeenkomst					
Overheidsinstanties Niet		9,17	32,550	-,640	,855
Inlichtingsdiensten Huisadres		9,42	26,869	,440	,817
Inlichtingsdiensten	Werkplek of	9,37	25,959	,609	,810
studieplek					
Inlichtingsdiensten		9,45	25,786	,710	,807
Uitgaansgelegenheid					
Inlichtingsdiensten		9,49	26,150	,661	,810
Ziekenhuisbezoek					
Inlichtingsdiensten	Politieke	9,47	26,178	,631	,810
bijeenkomst					
Inlichtingsdiensten Niet		9,14	32,443	-,625	,854