Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Dept. of Information Security and Communication Technology

Jonas Lillehovde

Security awareness and risk perception regarding data privacy of the digital natives

Master's thesis in Information Security
Supervisor: Vasileios Gkioulos & Gaute Wangen
June 2020

NTNU
Norwegian University of Science and Technology

Jonas Lillehovde

Security awareness and risk perception regarding data privacy of the digital natives

Master's thesis in Information Security
Supervisor: Vasileios Gkioulos & Gaute Wangen
June 2020

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Dept. of Information Security and Communication Technology

NTNU
Norwegian University of
Science and Technology

Abstract

To interpret if companies with different online business models sufficiently handle data privacy. It was beneficial to conduct a study towards the user's perspective, and the indicators that might affect behavior when users browse the internet. The sample chosen in this thesis consists of digital natives studying at NTNU and reside in Norway. First, to get an overview of what information is being gathered, a qualitative content analysis was conducted on the information in websites' privacy policies. This analysis was done on a smaller sample of websites with different business models and intentions, e.g., social media, online stores, and news sites. Furthermore, this thesis mapped the degree of awareness and risk acceptance in terms of data gathering online with a quantitative research survey. By analyzing the results based on indicators, the outcome produced an understanding of which potential indicators could affect the awareness and risk perception, as well as the digital natives' degree of security awareness and risk acceptance in regards to data privacy.

Sammendrag

For å tolke om selskaper med forskjellige forretningsmodeller håndterer personvern i tilstrekkelig grad, var det gunstig å gjennomføre en studie mot brukernes perspektiv og indikatorene som kan påvirke atferden til brukere når de surfer på internett. Utvalget i denne oppgaven består av digitale innfødte som studerer ved NTNU og er bosatt i Norge. For å få en oversikt over hvilken informasjon som samles inn, ble det først gjennomført en kvalitativ innholdsanalyse av informasjonen til nettstedenes personvern-policyer. Denne analysen ble utført på et mindre utvalg av nettsteder med forskjellige forretningsmodeller og intensjoner for datainnsamling, for eksempel sosiale medier, nettbutikker og nyhetssider. Videre ble graden av bevissthet og risikoakseptanse kartlagt med hensyn til datainnsamling ved en kvantitativ spørreundersøkelse. Ved å analysere resultatene basert på indikatorer, ga resultatene en forståelse av hvilke potensielle indikatorer som kan påvirke bevissthet og risikooppfattelse, som til sammen utgjør de digitale innfødtes grad av sikkerhetsbevissthet og risikooppfattelse når med hensyn til personvern.

Contents

Figures

Tables

Chapter 1

Introduction

1.1 Topics Covered

As technology becomes more and more integrated into our daily life, especially for digital natives, most of us interact with others connected to the internet every day through a variety of different apps, social networks, and apps. As a result, Internet privacy has gained increased focus the past decade, but how has this affected us as users? This thesis covers the topics of digital natives' security awareness and risk perception regarding data privacy when browsing the internet.

1.2 Keywords

Internet;data privacy;web sites;data collection;browsing;consumer privacy;privacy management;privacy policy;data security

1.3 Problem Description

In the wake of recent data scandals and the introduction of GDPR, digital natives are browsing the internet daily through a variety of devices. However, when interacting with websites, users are met with cookies and privacy policies that require their consent before accessing the desired content. Sometimes, it can feel like the current answers to the information gathered by online are hidden behind a labyrinth of clicks and a wall of text. With the current value of information in our modern society, users should be able to understand what they do agree upon before potentially accepting in ignorance. Having several laws preserving users' data privacy through non-functional protection mechanisms is undesirable, unwise, and an unwanted position for users on the internet. Creating potential exploitation, through companies profiting of user's information without their rightful knowledge. Major companies use this information to increase their income without the users being aware of who, what, and where the information about

them exists. This thesis examines what data websites collect and if there is a difference between the business models. Lastly, digital natives' security awareness and risk perception are examined to see if digital natives are aware of and accept risks that can threaten their privacy.

1.4 Justification, Motivation and Benefits

With the current development of technology and the impact it has on our daily life, users need to be aware of how this affects us. The internet helps us obtain, share, and manage information, but everything comes with a price. Companies use the information about users to personalize ads, target marketing strategies, and influence decisions, so the need for awareness is more significant than ever. Even though authorities such as the EU enforce laws provided to strengthen user's rights on the internet, this does not necessarily mean that the users are aware of their rights or the laws protecting them. This uncertainty is making the laws, therefore, not fulfill their intention. However, by raising the awareness of users when accepting the risks related to data privacy online, the users will be able to understand at least what information they are providing and what choices they have. Which, arguably should be viewed as the requirement for consent in the first place. Companies should be able to utilize users' information, but with consent where the user knows what information is provided and the purpose of the information. In general, this will benefit both companies and users, increasing awareness and making the use of information more legitimate and fair for both parties.

1.5 Research Questions

1.5.1 RQ1: How do information gathered differ depending on the business model of the website?

1.5.2 RQ2: To what degree are digital natives aware of the information gathering about their data when browsing the Internet?

Hypothesis 1: Higher education level achieved will result in an increased awareness for digital natives when browsing the internet.

Hypothesis 2: Digital Natives are more aware of information gathering when browsing the internet than non digital natives.

1.5.3 RQ3: To what degree do digital natives accept risk and willingness to provide information when browsing the Internet?

Hypothesis 1: Higher education level achieved decreases the willingness of digital natives to provide information.

Hypothesis 2: Digital Natives are less likely to accept risks than non digital natives

Chapter 2

Background

2.1 Privacy

Privacy is a universal term, with high importance to all kinds of people. People seek to protect the privacy that exists on a personal level for the individual and a business level for companies. Governments and institutions all over the world have tried to define privacy in order to enhance its importance. A quick internet search for the definition of privacy, returns thousands of results, and not a clear definition [1–3]. Privacy is a term discussed heavily in different philosophies across fields of work and aspects to life. People with backgrounds from the legal sector and sociologists have tried to define privacy over the years, with different aspects to the terminology.

Privacy must be clearly defined in order to create a foundation for what is meant when later discussing the understanding of people's awareness and risk perception, related to data privacy. Gavison, R. suggests in his paper [4] that the concept of privacy is *coherent and useful in three contexts: The losses of privacy, invasions of privacy and actionable violations of privacy.* Here privacy is mentioned in all of the three contexts, and each is a subset of the previous category, linking them together as a part of the same concept. Further, privacy is related to the concern of our accessibility to others. Having a perspective as privacy as a concern for limited accessibility enables the identification of when a loss of privacy occurs. Westin, A. [5] has defined privacy as *Claim of individuals, groups, or institutions to determine for themselves when, how and to what extent information about them is communicated to others*".

Further, privacy is considered as a vis-à-vis to others, meaning that privacy is a zero-relationship between individuals, persons to persons, or groups to groups. Therefore, for a breach of privacy to occur, the situation is dependent on intrusion from people on the outside. A person completely isolated without any connection to other people have zero risks of a privacy intrusion, before connecting to others in some way. Explicitly indicated by Shils, E. [6] that we consider the existence of

privacy only to exist within contexts that consist of interaction, communication, and perception. In a paper by Solove, D.J. [7], he argues that privacy is created by the society and that it cannot be understood independently from it. In a society, certain activities can be viewed as threatening to privacy. Solove describes these activities as *protection from a cluster of related activities that impinge upon people in related ways.* These activities are described as social friction, with privacy being the relief from the impingement they create. However, even if some activities are considered treating or problematic to someone's privacy, laws do not always account for every single case of these activities. For example, if consent is given, there is no privacy violation. Ultimately meaning that the law must be able to divorce each case individually, making privacy a problematic topic with defining what these threatening privacy activities are and if they apply for the specific context of the situation.

All of these different definitions and descriptions have some terms in common:

- The individuals rights
- Interference from unauthorized parts

Based on this previous definitions and the presented information about privacy, we can set a common ground and define privacy in this thesis as:

"*A protection or relief from activities in the society that threatens the control of own integrity and the freedom from unauthorized intrusion.*"

The first part of the definition a "protection or relief" describes privacy as a concept of limiting accessibility from others. "Control of own integrity" includes the claim that the individual should decide when, how and what information is available to others. "The freedom from unauthorized intrusion" reflects that privacy breaches can only happen if other people from the outside commit an activity without a given consent.

2.1.1 What is Data privacy?

Today, data is one of the most valuable assets for companies around the globe. Companies generate income through collecting, sharing, and using data, while users expect privacy and transparency on how the data is managed in return for their consent. Data privacy, also referred to as information privacy, serves as a branch of data security. These terms get mixed, but we can separate data security from data privacy by thinking about data security as something that protects data from compromise and attacks. In contrast, data privacy relates to governing how data is collected, shared, and used. Furthermore, none of the laws mentioned in this thesis concerning data privacy define a clear and precise definition of data privacy, making it just as "floating" as the privacy definition itself. Instead, they build an understanding based on best practices and explain the rights to users and companies to elaborate on what they mean by data privacy. Regarding these

best practices, several scientists and institutions have tried to come up with which practices account for data privacy.

In his paper: Resolving conflicting international data privacy rules in cyberspace, Reidenberg, J.R. [8] mentions a core set of fair information practices to assure that users (members of the society) understand and participate in the collection and use of their personal information. These practices revolve around four sets of standards: (1) Data quality, (2) Transparency or openness of processing, (3)Treatment of particularly sensitive data, and (4) Enforcement mechanisms. These elements ultimately describe the scope and values of data privacy. Further, the Federal Trade Commission [9] in the US made the *fair information practice principles (FIPPs)* which are guidelines to assure adequate information privacy protection. These guidelines consist of some core principles of data privacy: (1) Notice/awareness, (2) Choice/Consent, (3) Access/participation, (4) Integrity/Security, and (5) Enforcement/Redress. As can be seen, these principles also revolve around the same core aspects of data privacy, as previously mentioned, indicating that the principles have stayed stable over the years during the development of the technology. However, even though there is a theoretical understanding of what data privacy is and the principles data privacy revolves around. Data privacy faces some challenges when addressed in practice. These challenges arise because of the processing of data. It becomes a fine line of protecting individuals' privacy preferences and information, and at the same time, being able to use the data.

This leads to what data privacy accounts for, obviously data, but for users specifically one type of data: their personal identifiable data. Before the rise of computers, the collection of PII was less worrying. The only information about users was available to persons who had a relationship with the users in real life or by rumors by spreading the word around physically. Also, the information stored was limited and relied on papers, books, and manual record systems. When the computers were introduced, the information could now be collected and stored in quantities never imagined. Entities could now collect, organize, access, and search for data on a much larger scale than the manual systems [10]. This data collection can be associated with privacy expanding into data privacy. The introduction of technology and the rapid envelopment created a need for new concepts to ensure the protection of users. As with the privacy and the data privacy definitions, PII also seems to be a concept that is difficult to define clearly. PII has been defined in several ways, depending on the context. Both legal and technological scientists, as well as institutions, have written a variety of definitions. For example, the Data Protection Directive [11] defines personal data as: *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 identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity*. However, this is definition is meant to be very broad, while a more general way to put it would be:

Data are personal data, when the information in the data can be linked to a person.

Some examples of this kind of data are: insurance numbers, email addresses, phone numbers, IP-addresses, geolocations, and more.

2.1.2 Why raise awareness?

As most of the world is online, enjoying connectivity and interactions, the internet provides us, and more and more are starting to highlight the most prominent companies' ethics and morale in their service. Over half the world's population have been interacting on the internet, and it has become quite abnormal not to use any form of online services or devices. Amnesty International describes the issue in their paper about Surveillance Giants [12]:

> *Every time we interact with the online world, we leave behind a data trace, a digital record of our activity. When we send an email, the content of the message, the time it was sent, who it was sent to, from where, and a host of other information, is recorded and stored in servers and data centres. A similar process happens when we browse the internet, use an app on our phone, or buy something with a credit card.*

Etzioni, A. describes privacy as merely *a good among other goods and should be weighed as such.*[13]. As the technology is developing fast, the concern of privacy also increases proportionately as a direct variation. Several papers and articles express that technology has changed our attitude and beliefs to privacy. [14–19]. To fully understand why scientists and journalists try to raise awareness and warn people about their privacy concerns, it is needed to see what harms the breach of privacy can cause to the individual as well as the society. First of all, according to the definition from Merriam-Webster[20], "harmful" is considered *of a kind likely to be damaging*. Privacy breaches can trigger harmful effects on individuals and can cause consequences for both individuals and companies. Furthermore, a look at specific harmful effects from previous research is necessary to understand the effect breach of privacy can have on members of society. Van den Hoven, et al. [21] explains 4 moral reasons for protecting our data in their paper about data privacy, the list is inspired by a lecture in the subject data privacy by Vinterbo, S.[22]:

1. **Prevention of harm**:
   Non authorized persons getting unrestricted access to, for example, accounts, profiles, repositories, etc. can be used to cause harm towards the data subject. Example: identity theft and fraud.
2. **Avoiding informational inequality**:
   Companies use of users' data is something the average user has little to zero control over. It is somewhat controlled and regulated under data protection

laws, but the big companies control most of the market. These companies can present data to users by direct marketing, micro-targeting, and demanding consent for access, which ultimately creates an imbalance in the power relationship between users and companies. Example: informational service providers.

3. **Preventing unjustice and discrimination**:
   Context integrity is a major concern concerning users PII. Some information can change other's perceptions and create disadvantages or discrimination if taken out of context and used in another sphere. Example: health information and insurance
4. **Encroachment on moral autonomy and human dignity**:
   Without privacy, people in a society would be under a constant exposure of influences and moral judgment. Leading to a change of behavior and making decisions they would not otherwise have made. This change can create a chilling effect on both the individuals and society, which violates humans' right to freedom and respect to individuals' dignity. Example: manipulation of social media for political objectives and mass surveillance.

2.2 Terms and descriptions

2.2.1 The digital natives

The term digital natives are used to describe people born and bred after the social-digital technologies became available. In the paper, understanding the first generation of digital natives by [23], digital natives are described as people born after 1980 and have the skill to use technology. 1980 marks the transition for when technology was used to communicate with each other. The use of computers to share documents, and later e-mails were adapted even before the introduction of the world wide web in 1991. Several definitions of the term digital natives exist, some more diffuse than others. However, what they have in common is that it involves the people born during the digital era. In this thesis, digital natives will follow the description above, defining digital natives as *people born after 1980 born in the digital era.*

2.2.2 Privacy related terms

When browsing the internet, embedded features exist on websites all over the world wide web to improve and enhance our browsing experience. These are often used to enable personalization by remembering choices for a limited time or until deleted by the user. While this might increase efficiency and enhance the user experience, these methods are also used to collect data about users for other purposes. In this sub-chapter, some of the most regularly and currently used features are presented to create an understanding of which data collection features exists, and the current state of the art of these methods.

Cookies

A cookie file is a text file stored in the browser's folder or subfolder. These files are created by the web pages a user accesses and are accepted and processed by the computer's browser software. Cookies are used to remember information for users on their computers, so the next time a website is visited, it will remember preferences and choices. Furthermore, cookies can contain every kind of information, for example: time of visit, items added to the basket, all links clicked. Originally, they could only contain a limited amount of text, and the size was limited. However, websites developed third-party cookies. These cookies stores a unique ID on users' computer, while the rest of the data is stored on their systems. Resulting in that the websites with third-party cookies can recognize users and access this stored information. By having bits of other websites embedded onto the original website, other websites can identify the users to track their activity and personalize ads towards them.

Cookie definitions: These definitions are based on the ICO guidance [24]

- First and third-party cookies:
  First party cookies are cookies set by the current website the user is visiting. On the other hand, third-party cookies are set by a domain other than the website the user is being visited.
- Persistent cookies:
  Remains on the user's device for a stated period of time. They are activated each time the user interacts with the same website that created the cookie.
- Session cookies:
  The session cookies refer to temporary cookies that allow website operators to link user actions during a browser session. The session starts when the browser is opened and ends when the user closes the browser. When the browser is closed, all session cookies are deleted.

Cookies can be divided into different categories of cookies, note that cookies may function in more than one category.

- Strictly necessary cookies:
  Essential to move around and use features on the web site, for example, putting items into a shopping cart. Automatically set when the web page load and the web page would not function as intended without them enabled.
- Performance cookies:
  Used to improve performance on the website. Examples are error management, testing designs, and analytics. However, these cookies are not used to re-target adverts but serve as a tool for improving the web site.
- Functionality cookies:
  Often used to remember the result of a user action, for example, remembering settings, not offering the same service again, and remembering choices. Functionality cookies can also be implemented for services offered to the

users without any request, such as a survey or feedback.
- Targeting/advertising cookies:
  Generally, third-party cookies contain a unique key used to distinguish individual users. Third-party organizations can place these cookies with permission from the website owner to collect browsing habits and preferences. The information gathered is used to, for example, target adverts to the user or gather information about the effectiveness of an advertising campaign.

Data broker

In the days of big data, user's data and personal preference are monetized and traded between companies. The value of data has only increased with the years surpassing other resources like oil, arguing, making data the most valuable resource in the modern world according to some sources [25]. Data brokers are companies that collect consumers' personal data to sell with other companies. These transactions often happen without the consumers' knowledge or consent, taking place in the shadows. By aggregating raw pieces of individual information, these data brokers can compromise the user's right to privacy [26] Especially, policymakers the last decades have raised concerns regarding the lack of transparency of data brokers [27].

Web beacons

Web beacons, also known as web bugs, pixel tags or clear gifs, is a type of embedded content where the *content itself is irrelevant, but the request for content carries useful information* [28]. These web beacons are often used together with cookies in order to monitor user's actions. Web beacons are placed in the code of, for example, a web site to see the site visitors' behavior. When the code is invoked, it will simultaneously transfer information such as IP addresses, timestamps for when, and for how long the web beacon was viewed [29].

2.3 Laws and Regulations

As previously mentioned, the term data privacy has seen a variety of different definitions. Moreover, it has also been a subject in several laws over the past decades. In this subchapter, a timeline is presented for the laws in Norway regarding data privacy to create an overview before some further elaboration is conducted on the current laws related to data privacy for the citizens of Norway.

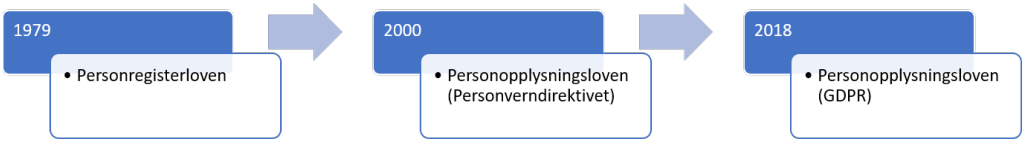


Figure 2.1: Timeline of privacy related laws in Norway

2.3.1 Personopplysningsloven

In Norway, the primary law regarding the processing of personal data is "Lov om behandling av personopplysninger (personopplysningsloven)". The the purpose of this law is to protect individuals against privacy violations through the processing of personal information. Personopplysningsloven replaced the previous law with the same name on the 20th of July, 2010, which had the same title. Figure 2.1 shows the timeline of the previous laws regarding the processing of personal information in Norway. This new law, from 2018, implements Eu's GDPR as a current law in Norway. Further, the new law now accounts for all processing of personal information, compared to the prior version that only covered systematic storing and compiling of personal information. As mentioned, Personopplysningsloven implemented GDPR into the Norwegian legal system. In chapter 4, some exceptions from GDPR in Norway can be found regarding the right of access and obligations to inform, continuing the prevailing law implemented before GDPR.

2.3.2 GDPR

Even though Norway is not a member of the EU, Norway is a member of the European Economic Area. As a result, Norway is bound by the GDPR in the same manner as the EU members. GDPR was implemented in the EU on the 25th of May 2018, but because of delays the law was not implemented in Norway before the 20th of July. Ultimately, GDPR implements stricter rules on data protection, which means primary two things as described by the European Commission [30]:

- People have more control over their personal data
- Businesses benefit form a level playing field

In practice, companies receive more responsibility in terms of overview and control of what personal data they collect, store, and process. This responsibility requires the companies to implement routines before, under, and after they process data. Furthermore, companies must also be able to document that they are acting according to the law, or they can face sanctions.

For individuals, GDPR is designed to help protect the rights of individuals. The law introduces 8 rights, as can be seen in the list below. The rights can be found in the regulation itself or as a summary from data privacy actors [31, 32].

- The right to be informed
- The right of access
- The right of rectification
- The right to erasure
- The right to restrict processing
- The right to data portability
- The right to object
- Rights in relation to automated decision making and profiling

2.4 Methodology background

This chapter contains information concerning methodology. The chapter describe theories for structuring research, philosophy of research and provides a foundation for the methodology chapter where the approach for this thesis is described.

2.4.1 Background

In order to find a suitable method, different methods were considered up against each other, providing the most optimal method for studying the target group. The optimal method can be found by a discussion regarding the choice of the philosophy of science, the research design with the characteristics and purposes associated with each design, the data collection, and the data analysis. The method describes how we establish reliable and durable knowledge through the thesis's lifespan, providing validity to the methods used for collecting data. The structure of this chapter is based on the research onion by [33], which illustrates the connection of choices between methods and their respective design and characteristics.

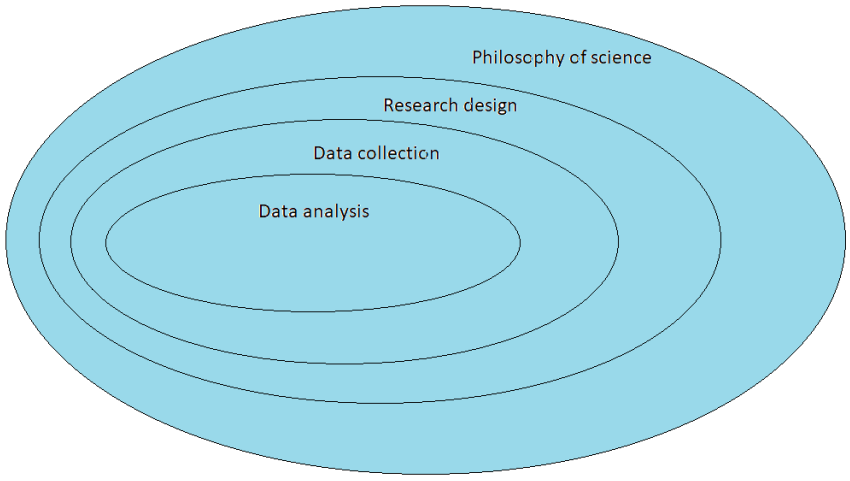


Figure 2.2: The research onion [33]

2.4.2 Philosophy of science

Practical Research: *planning and design*[34], gives two examples of general assumptions that underlie many research studies:

1. The phenomenon under investigation is somewhat lawful and predictable; it is not comprised of entirely random events

2. Cause-and-effect relationships can account for specific patterns observed in the phenomenon

To justify these assumptions that underlie research studies, distinguishing between different philosophical orientations is valuable—I.e. theories for the creation and understanding of science. Two terms associated with philosophical orientations are: (1) Positivism and (2) Constructivism. Positivists are under the assumption that with appropriate measurement tools, *scientits can objectively uncover absolute, undeniable truths about cause-and-effect relationships within the physical world and human experience* [34]. On the other hand, constructivism views the world independent of the human mind and has abandoned the idea that absolute truths are found in the natural world. Furthermore, constructivism emphasis on subjectivity and bias, rather than objectivity with scientific approximations.

2.4.3 Research design

Primarily, a research design will vary between either a qualitative design or a quantitative design. Qualitative methods collect data by examining in-depth and study phenomenons and existing events. Qualitative data can be collected from various sources, such as interviews, observations, or questionnaires. However, in qualitative designs, the sample is not drawn; the participants are either enlisted or recruited by the researchers. After the data is collected and ready to be analyzed and compared, in-depth answers, and a holistic understanding of the phenomenon can be examined [33].

On the other hand, quantitative methods involve collecting data from larger samples with comparable data. The research questions are exact and easier to measure, which divides quantitative methods form qualitative methods. Measuring, counting, and using statistics are central in quantitative methods, where the researcher tries to establish the scope and get an overview of the context of the data [33].

In some cases, a combination of the two designs is used, known as mixed methods. Where often, the qualitative method is used for the in-depth examination. It allows the creation of hypotheses before quantifying the data and using a quantitative method to test the hypothesis [35].

2.4.4 Data collection

Collected data can be a source of primary data or secondary data, collected through a qualitative design, a quantitative design, or mixed methods. Quantitative designs contain easy and effective data gathering methods, but with less possibility for complex and advanced analysis. A qualitative design is better suited for advanced analysis, as seen in complex and more undefined research questions [36].

In a cross-sectional study, all data will be collected at once. Collecting data at

once makes it easier to conduct than longitudinal studies. On the other hand, longitudinal studies follow a single group over several months or years and collect various data [34].

2.4.5 Data analysis

The primary purpose of the data analysis is to process the collected data to present the most important information related to the results. This is done by processing the qualitative data into measurable data and quantify the answers from, for example, questionnaires to raw data used to produce statistics [34].

Chapter 3

Related work

The related work chapter will be divided into a section for each research question. Related work conducted by other authors will be discussed within the scope of the research questions of this thesis. The outcome will contribute information about the state-of-art, as well as demonstrate the methods used to approach similar research questions in other studies.

3.1 How do information gathered on a selection of websites differ depending on the business model?

Personalization and the gathering of user's information are seen hand-in-hand on the Internet today. Websites use cookies to remember information about users when visiting websites. With the introduction of third-party cookies, websites now store data on their systems and threaten user's privacy rights online. Today, almost all websites use cookies. Hoofnagle and Good [37], shows that 87% of websites from a top 25000 list used cookies to store information. To collect information about what types of cookies and the different elements of data they collect, one paper from Cahn, A., et al. [38] used Cookiepedia to analyze cookies. Cookiepedia [39] state that they are the largest database of pre-categories cookies and online tracking technologies. The database is maintained by a privacy management software company called OneTrust. Cookiepedia is used to enlighten users about what cookies do, who is using them, for what purpose, and lastly how to manage them. By searching for a specific website, the database will look up the cookies on the website and provide a classification quantification of the amount found.

On the other hand, avoiding these companies from the gathering of information about users on websites are difficult. The Soltani, A., et al. [40] conducted a study in 2009 and found that the top 100 websites from a selected list gathers information about users, and indicated that they even use techniques to re-instantiate deleted cookies. However, this was the case of flash cookies, which according to the same paper, 50% of the websites used. As a follow up by the same authors in

2011, Ayenson, Mika D., et al. [41], concluded that the problem still exists and that users have a hard time avoiding tracking. Too notify users about the gathering of their information and require consent, GDPR was implemented in 2018 to increase the user's control of internet privacy. Dabrowski, A., et al. [42] shows that after the implementation of GDPR, that EU consumers encounter significantly less unconditional usage of persistent cookies. Suggesting that some changes have happened to the content of the policies and the presentation of cookies. Laws in EU regulate cookies, Proton Technologies, a co-founded project of the Horizon 2020 Framework Programme of the European Union [43], showed how the ePrivacy Directive and GDPR requires websites to fulfill a list of requirements in order to become compliant with the regulation. However, being compliant does not necessarily mean that users understand what the policy express through the cookie design. To analyze the content of the cookie policies, several authors have [44–46] used content analysis. Anton, A. I., et al. [45] used goal-mining to measure if the website's operations contradict or do not fulfill the requirements made by the privacy policy. They measured the requirements by using a privacy goal taxonomy based on protection goals and vulnerability goals.

Further, Earp, J. et al. [44] made a list of 24 selected websites from a variety of industries. Where they connected the results from the policy to a taxonomy, providing an overview of the frequency of occurrences within a given category. The frequency combined with a survey gave results that indicated a gap between what users value and what website privacy policies emphasize. A different paper by Pollach, I. [46], conducted a content analysis combined with a critical linguistics method to identify weaknesses and make suggestions for improvement. Based on a sample of 50 web sites with 4 different business models, the paper suggested that online privacy policies are more written in terms of avoiding litigation, rather than raising user awareness. Further, the paper expresses the need for a need for changes in the presentation format of privacy policies. With that in mind, Miyazaki, A. D. [47] suggests several recommendations for public policy makers in order to improve the user's understanding.

Moreover, these policies need changes in content but also in the presentation format. McDonald, A. M. et al. [48] suggests in the discussion part that it is not the policy format itself that is confusing; rather, it is the reader's understanding of where to find information. Where, Kelley, P. et al. [49] explored a solution for making a single page summary of the policy based on a grid design, which made it more "pleasurable". Earp, J. B., et al. [50] addressed the need for more comprehensible and concise privacy policies online. By designing three types of alternative privacy policy representations, the paper compared the consumer perception and comprehension of the typical online privacy policies versus their alternative ones. Results indicate that the typical online privacy policy seen at most websites are the least comprehensive and does not appeal to the consumers.

3.2 To what degree are digital natives aware of the information gathering about their data when browsing the internet?

Various studies over the past decades have shown that users are concerned about their online privacy [51, 52]. However, users seem to forget about their concerns regarding privacy online and prioritize access to content even when the most personal details are communicated with any compelling reason to do so [53]. To measure the degree of awareness, different methods have been seen used. In a paper by Hoebel, N. and Zumstein, D., [54], they conducted a quantitative survey towards a university sample with respondents younger than 35 years. The result of the survey indicates that respondents feel monitored while surfing the internet, and at the same time, do not like to reveal personal data online. Gerber, N. et al. [55] examines users awareness about information gathering through a qualitative method of data collection, more specifically, a study with semi-structured interviews with 24 participants with a different background. Furthermore, they raise a question about whether quantitative methods are suited for data collection in regards to user awareness and risk perception. As a result, a discussion of what method is best suited will be discussed in this thesis.

On the other hand, Gunleifsen, H. et al. [56] used an online survey to identify the general stance towards IT, knowledge, risk evaluations, and trust in authorities. The sample who responded had an average age of 56 years and primarily resided in rural Norway. Another report by Ariu, D. et al. called the security of the digital natives [57], studied the level of awareness and perception of IT among 1012 university students in Italy. The study conducted 60 multiple-choice questions related to different aspects of security awareness, mainly towards IT security issues in mobile devices but also towards an approach to internet use, passwords and risk perception. The findings from this study indicate that the digital natives have a wrong perception of their knowledge and awareness of information security. Digital natives also lacking awareness of protection methods, and tend to choose usability over security.

As a follow up on this study, Gkioulos, V. et. al. [58] identified how user confidence, security awareness, and background affects digital natives mobile decisions related to security impact through a survey. Furthermore, the paper also divided the sample into groups based on security competence (Generic, medium and high-security competency), as well as other indicators used for evaluating security awareness. The results from this follow-up study indicate that specific areas of user behaviour of digital natives are not significantly affected by their security awareness or background. However, results from the study indicate that higher awareness in terms of security risks leads to more willingness to opt for security when practical solutions become available.

3.3 To what degree do digital natives accept risk and provide information?

Several papers and reports have been investigating risk perception towards internet activities. Forsythe, S. M. and Shi, B. [59] examined risk perception of internet users through a framework of risk perception. Fist, the paper identified potential risks before a survey was sent to a total of 641 respondents. The results from the study on how users perceived risks and were placed into a framework.

A series of reports from NorSIS [60–63] called "the Norwegian Security Culture" has conducted a series of studies towards Norwegians with different questions related to risk perception. Through research study, Norsis measured how much risk on a scale from 1-5, where 1 is "not worried at all" and 5 is "significantly worried". Which is interesting, considering this thesis also was looking to conduct the same method. Furthermore, the reports claim that too few Norwegians are given cybersecurity education and that the current education does not have sufficient effect. The latest from the same series of reports from 2019 [63], examines the risk perception of the Norwegian society by a research survey. Results show an increased fear when interacting with certain services online. The same report also shows a slight increase in cybersecurity training compared to the results of the previous reports. However, even with the increase, the report indicates that the number of people with cybersecurity training was not sufficient.

Furthermore, Lynne, M. et al. [64] explored the framing and personality factors that affect privacy-related decision making. The study measured the acceptance of cookies through a qualitative method. An older study from Adams, A. [65] found that there was a mismatch between the users' perception of privacy risks and their realization of actual privacy risks. To measure risk perception, Bhatia, J. et al. [66] introduced a theory of vagueness for privacy policy statements based on a taxonomy of vague terms from an empirical content analysis. Further, the paper indicated that vagueness in privacy policies could introduce privacy risk by concealing privacy-threatening practices behind vague terms and unclear sentences. The results lead towards that users are accepting risks and providing information without an understanding of the consequences of giving their consent. To measure willingness, the paper created scenarios with benefits and risks connected to a statement in the scenario. The respondents had to answer with a scale of degree of willingness from "extremely willing", ranging to "extremely unwilling". Factorial survey methods such as the use of scenarios are valuable for studying factors related to the perceptions of the respondents.

A publication from Sage research methods [67], explains how it can be desirable to gain more in-depth insight into decisions made when responding to questionnaires than using only single-item questions. By using vignettes, detailed descriptions of situations make the respondent's judge stimuli and help the researcher

get a deeper insight into the respondents' judgement principles. This approach was used by Hibshi, H. et al. [68] in their paper regarding the assessment of risk perception in security requirements composition. Here, the vignettes used in the survey were designed with dimensions that influence the perceived level of security risk.

Results from the follow-up study by Gkioulos, V. et al. [58] show that digital natives are willing to accept risks despite their concerns about security. However, there was no significant effect by the overall knowledge about security.

Chapter 4

Methodology

4.1 Choice of Methods

This chapter contains a discussion about applicable research designs and data collection methods to answer previously stated research questions. As a result of the discussion, the final choice of methods and the background of choosing exactly these methods are provided. Lastly, this chapter also describes selecting, approaching, and receiving responses from the target group.

4.1.1 Philosophical orientation

The philosophical orientation in this thesis was conducted from a philosophical theory perspective called positivism. However, the philosophical perspective had a slight variation of positivism, called post-positivism. This perspective is less self-assured and more tentative, while at the same time having an objective view. Furthermore, post-positivism has an understanding that the potential conclusions to these research questions can not guarantee and define the absolute truth, which refuses the constructivism's view towards science. Post-positivism was ultimately chosen because a solely perspective as positivism would assume that the conclusion leads to proven results. The results of this thesis should be viewed as an increased probability that such-and-such is true. [33]

4.1.2 Research Design

In research design, the main reason is to lay a foundation for how to collect data. To answer the research questions regarding digital natives and if certain factors affect their awareness when browsing the internet, specific designs would be too time-consuming for this thesis In order to get a good representation of the digital natives as a population, a design with a larger data set is preferable. Furthermore, as a result of choosing a larger data set, the quantitative design is well-fitted for the propose of this thesis. Allowing effective data collection through time-efficient methods increases the amount of data for analysis and the possibility for results

to answer research questions. Moreover, this provides a higher validity, reliability, and generalization to the digital natives as a population. However, to analyze the content of websites, a more in-depth focus with smaller sample size is required, resulting in the most suited form of data collection is through a descriptive research design. Descriptive research "*involves either identifying the characteristics of an observed phenomenon or exploring possible associations among two or more phenomena*" [34].Moreover, descriptive research examines a situation "as it is", without modifying or changing the variables. Some commonly used methods for descriptive research designs are: observation studies, correctional research, developmental design, and survey research. Lastly, a descriptive research design is suited for a design with a one-time data collection.

4.1.3 Data Collection

Primary data collected in this thesis was based on both a quantitative design and a qualitative design. The combination of methods is referred to as a mixed-methods design, where both methods were conducted throughout the thesis. The reason being, the scope of this thesis was limited to one semester regarding its duration. Limited time was rejecting the opportunity to collect data over a more extended period, while also having enough time to analyze it. The data collection method most suited for this thesis was, therefore, a cross-sectional study. Since time was a limited resource, the longitudinal study was found not suitable for the scope of this thesis.

For the qualitative design, the method used to obtain data was through content analysis. Content analysis is a research method conducted on texts in e.g., documents, to analyze patterns through *systematic examination of communicative material*[69]. Furthermore, by conducting this analysis on a smaller sample of websites. A structured and categorized overview of the content the policies contain was presented.

Moreover, another method used to obtain data was through survey research. Because a potentially well representative sample of digital natives is studying at NTNU, access to students and employees within the correct age group was reliable for recruiting respondents. Combined with the fact that students can be approached on campus, results in a questionnaire chosen as the primary method for data collection. Questionnaires through survey research include acquiring information about, e.g., characteristics, opinions, attitudes, experience, etc. about one or more groups of people and tabulating their answers, as seen in Practical Research: *planning and design*[34]. In this method, a series of questions are conducted on willing participants. However, a potential risk is that questionnaires require awareness of possible mistakes, such as generalization and validation, which will be further elaborated in the risk chapter.

For secondary data, information/document analysis, and the internet as a source will be gathered as supplementary information to the primary data collected. This secondary data was based on information produced by other people, which was produced for other projects. However, the data can still be valid and referred to by having a critical view of the usefulness and transferability of the data.

4.2 Data Collection

The data collection chapter describes the different methods for empirical data collection conducted. This includes background for the method chosen, the planned method, the conducted method, and a reflection of the method. An illustration of the data collection method can be seen in the figure below:

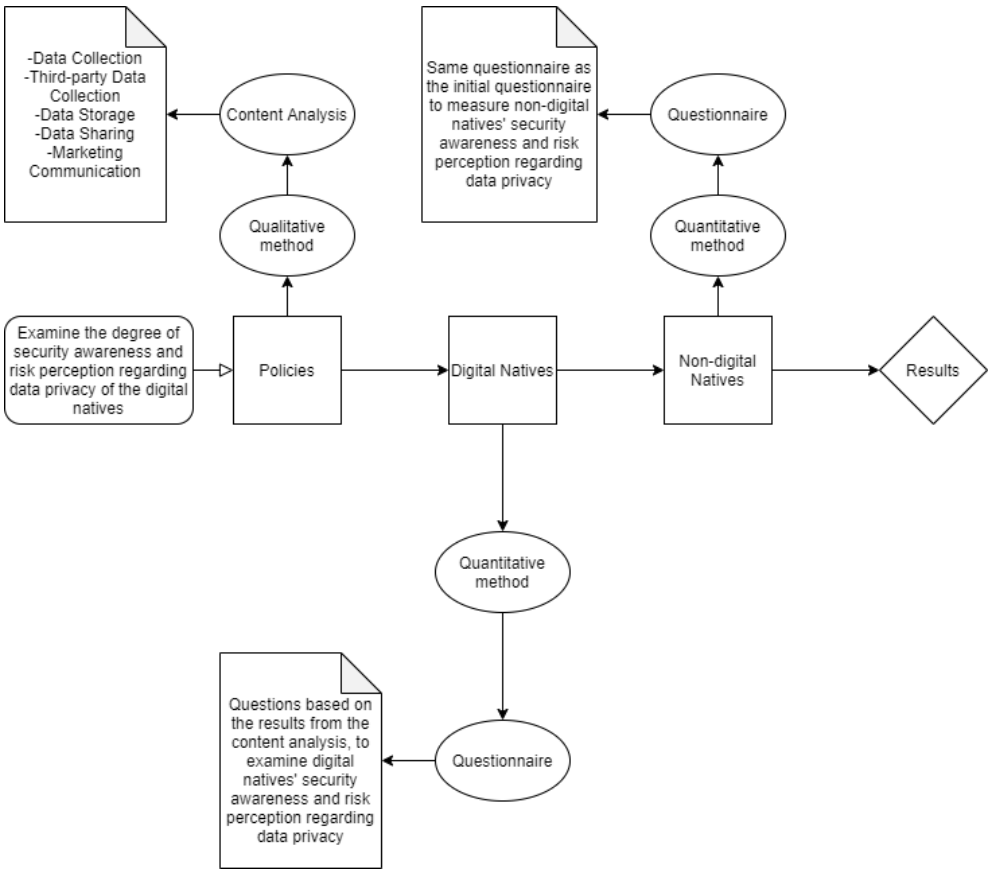


Figure 4.1: The process of data collection methods illustrated in a process map

4.2.1 Content Analysis

Background

To understand what type of information the different websites on the internet collect, store, and transmit about their users, a method of acquiring this data was needed. Research question 1: *How do information gathered differ depending on the business model of the website?*, was made to examine the gathering across multiple websites. Several business models exist online, and the way they make their income varies from the website's concept. For example, social media makes money selling the possibility to target their users with advertisements, while a retail or e-commercial store provides goods to be bought online. With this different background, one would expect slightly different ways to handle information.

Furthermore, there were two ways of approaching the research question, the first option being a content analysis of what the different websites state that they gather in their policy. This provides an easy and reliable way to access data since the websites are required to have this information available. Furthermore, it also provides repeatability since the same information can be looked up within all the policies. However, what a policy state and what actions the company conducts, can not directly be measured through reading the policy. The other option was a technical test to measure what the websites collect, store, and transmit [70]. A more technical method requires more time and more in-depth technical insight than content analysis. Nonetheless, it assures a more secure way of knowing what and how the companies behind the websites process information. However, with the resources and the limited time of this thesis, the primary method to examine research question 1 resulted in content analysis.

The content analysis was conducted to build a foundation for the following questionnaire. By understanding the types of information the different business models gather, using these results, realistic scenarios in the questionnaire could be created. However, since the results from the conducted content analysis were not intended to describe and find weaknesses in the different policies. The method does not differentiate on how the policy is structured (length, sections, phrasing, etc.); it simply relied on the data the policies contain.

Conducted method

Content analysis was conducted on a selection of websites. This selection of websites was based on recent statistics from various sources, depending on the specific type of content provided on the web page. For example, news sites were based on reported readers, while social media's popularity was based upon the number of users registered or active members [71]. Table 4.1 below showcase the websites visited for content analysis.

Nr.	Newssite	Social Media	E-commerce	Blog
1.	VG	Facebook	Komplett	Blogg.no
2.	Dagbladet	Snapchat	Zalando	Blogger
3.	Aftenposten	Instagram	Elkjøpt	Squarespace
4.	Nettavisen	Linkedin	Ebay	Wordpress.com
5.	DN	Twitter	Ikea	Wordpress.org

Table 4.1: Selection of websites chosen for the content analysis

To answer research question 1: *How do information gathered differ depending on the business model of the website?* a method based on the paper by Pollach, I. [46] was conducted. Pollach's method includes answering questions regarding key privacy concerns of Internet users within 5 different categories: data collection, data storage, data sharing, third-party data collection, and marketing communication. Each category contains questions to ensure that the coverage of policy statements according to GDPR and other obligations. However, since the paper's method was dated back to 2007, some slight changes were made to include recent changes relevant to current policies. Changes made to laws and the introduction of GDPR require websites in the EU to comply with the law or face charges in terms of fines. Some of the changes made to the original questions can be found in the GDPR legislation as seen in Personopplysningsloven[72]: the right to insight (art. 15 GDPR), the right of deletion (Art. 17 DSGVO), the right to correction (art. 16 GDPR), the right to withdraw consent (Art. 7 GDPR). See the table 4.2 below for changes made to the original method from Pollach, I.

	Changes made to the method
1.	Added “Geolocations” to Data Collection category.
2.	“Types of data collected by third parties” in Third-Party data collection category changed to “Data collected by third parties”.
3.	Changed “unauthorized employee access” to “unauthorized access” in Data Storage category.
4.	Added “User’s right of access to information” to Data Storage category
5.	Added “User’s right of erasure” to Data Storage category
6.	Added “User’s right of withdrawal of consent” to Data Storage category
7.	Added “User’s right to rectification” to Data Storage category
8.	Added “Informs about the duration of data stored” to Data Storage category
9.	Removed question regarding “business agents” and replaced with “third-parties” in the Data sharing category
10.	Changed “selling of data” to “refrains from selling of data” in the Data Sharing category to easier answer the question in a yes or no fashion.
11.	Added “Newsletters” to Marketing Communication

Table 4.2: Changes made to the original method from Pollach, I.[46]

With these changes made, the method becomes more relevant and applicable to today’s technology and laws. The five categories remain the same, while some of the questions got renewed. The table below provides an overview of the categories, together with the updated questions.

Category	Questions
Data Collection	Collection and storage of PII; Collection of aggregate information; Users' ability to view and update data profiles; Collection of user data via surveys; Sweepstakes used to gather customer data; Obtaining user information from other sources; Storage and usage of email addresses from inquiries; Cookies; Information on disablement of cookies; Information on consequences of disabling cookies; Web beacons; Geolocation;
Third-Party Data Collection	Types of data collected by third parties; Third-party cookies or Web beacons; Privacy agreement with third parties collecting data; Opt-out of Third-party data collection;
Data Storage	Measures taken to ensure secure offline storage of data; Measures taken to prevent unauthorized access; User's right to access information; User's right to erasure; User's right to withdrawal of consent; User's right to rectification; ability to delete PII; Records of PII kept after user deletes PII; Informs about the duration of data stored;
Data sharing	Sharing of PII with affiliates; Sharing of aggregate information with affiliates; Sharing of aggregate information with third parties; Sharing of PII with third parties; Refrains from selling of data; Sharing of email addresses; Sharing of data obtained in sweepstakes/surveys;
Marketing Communication	Unsolicited email; Unsolicited email from third parties; Newsletters;

The results from reading through the policies resulted in an answer within 3 results: "Yes", "No" or "Dno" (short for Do not know). For example, if a policy states "If you delete your account, we will erase all data stored about you", this means that the question "records of PII kept after the user deletes PII" is answered with "Yes". On the other hand, if an answer is not answered, the answer would result in a "Dno". As a result of this, to answer a question with "no", the policy would

have to state explicitly that the website does not conduct the activity stated in the question. When answering these questions while reading the policies, the “at-least-some” rule was applied. The "at-least-some" rule means that the practice would be considered true, even if the statement was not written exactly

4.2.2 Questionnaires

Background

In addition to the content analysis, a questionnaire was conducted to examine the security awareness and risk perception for digital natives concerning data privacy. The questionnaire was the primary data collection method and was built upon the data collected from the previously conducted content analysis. The results of the questionnaire contribute to answering research question 2: *To what degree are digital natives aware of the information gathering about their data when browsing the internet* and research question 3: *To what degree do digital natives accept risk and provide information*.

The target group for the questionnaire was digital natives associated with universities in Norway. Mainly, NTNU: Norges Teknisk-Naturvitenskapelige Universitet, which includes students and employees from all over Norway. There were several reasons for choosing a university as a sample. First of all, to maintain sample control, people associated with NTNU receive a FEIDE user. This kind of user is only provided to students or employees at universities in Norway and assures a secure login portal within education and research. Having a FEIDE-user in practice is making sure that the respondents of the questionnaire are authenticated in a matter that ensures that they are within the target group. Furthermore, this authentication ensures that the respondents are only able to answer once, preventing the forging of answers, and ultimately increasing the validity of the answers. Moreover, the recruitment of respondents was one of the most crucial parts of the research to succeed with surveys and questionnaires. Therefore, to ensure enough answers from respondents for the data analysis, the choice to recruit people associated with education and research appeared to be a valid method of reaching out to digital natives.

To create the questionnaire and gather responses, Nettskjema is a tool available for Norwegian students. This tool provides some benefits concerning data storage, as well as functionality and sample control. Nettskjema provides the FEIDE login option. As previously mentioned, by relying on FEIDE login, control of the respondents belonging to the intended target group is in place. Furthermore, in terms of data storage, Nettskjema has been validated and approved by NSD: Norsk Senter for Forskningsdata and REK: Regionale Etiske Komiteer for helseforskning for data collection and storage.

The project was reported to NSD to assure that the data collected did not include

personal identifiable data, the need for consent, and the anonymous data collected. This questionnaire received approval from NSD, making it legitimate for data collection. It did not interfere with users' privacy or collect personal identifiable data, as well as voluntary participation.

Conducted method

By conducting the questionnaire, the intention was to examine research questions 1 and 2, regarding security awareness and risk perception. To answer these research questions regarding security awareness and risk perception, the questionnaire was divided into three sections: 1. Background information about the respondents, 2. Security awareness, and 3. Scenarios for risk perception. The first part of the questionnaire mainly collects background information regarding the demographics of the users (age and gender). Since the sample for this questionnaire was NTNU students, it was interesting to know what faculty each respondent belonged to, as well as the highest degree of education level. Lastly, as a more privacy-related question, the respondents were asked to reply to an estimate of how many hours they surf the web each day and which web browser they primarily use.

The background information and demographics laid the foundation for measuring differences between groups later on in the analysis. Further, the second part of the questionnaire collects information to measure security awareness among the respondents. In this section of the questionnaire, respondents were asked about their knowledge regarding terms related to data privacy. The knowledge regarding the terms, helped identify the degree of security awareness. Security awareness was to be determined by the knowledge of cookies, data brokers, web beacons, and users' rights regarding GDPR. The four first questions asked about the respondents' knowledge on a scale from "No, I was not aware" - "Have only heard about it" - "Yes, I am partly aware" - "Yes, I am fully aware". Under the question as a description, a definition/explanation of the term was provided to remove the possibility that the respondent might think he/she knew the answer when this was not the case. For example, if a respondent had seen cookies on websites and believed that they are used in terms of privacy choices, without knowing that it is a file stored on their computer. Their answer would change from "Yes, I am fully aware" till "Yes, I am partly aware" after reading the definition provided.

The last part of the questionnaire consisted of questions to measure risk perception among digital natives. This part of the questionnaire was split into three different techniques for measuring risk perception. The first part asked, "what action do you often do when faced with cookies", with a picture that showed the options a standard cookie provides users. When interacting with the cookie choices, one can decide how much information the cookie can store. As described in the background chapter under the subchapter "Cookies", cookies can be divided into different cat-

egories. If the user wanted to limit the information gathered, one could decide to reject selected categories of cookies. To answer the question, the respondents had the choice of "Accept all" - "Remove targeting/advertising cookies" - "Remove functionality cookies" - "Remove all, but strictly necessary cookies". Moving on, a question regarding how many read the policies were asked the respondents, "Do you read privacy polices before accessing a website?" with the possible choices of "Yes, always" - "Yes, sometimes" - "No, never".

The next way to measure risk perception in the questionnaire was done through scenarios. The respondent would measure the risk of realistic scenarios by providing answers to scenarios based on the information found in the content analysis. This method was inspired by a paper written by the authors Bhatia, J., et. al. [66]. These scenarios present a description of scenarios with the benefit of using the service, as well as the risk involved. To answer these scenarios, the respondents could choose between four different options: "Very unwilling" - "Unwilling" - "Willing"- "Very willing".

The last way to measure risk, and also the last question of the questionnaire was given with checklist alternatives. Here the respondents had to make a choice regarding "have you done any of the following measures to protect your privacy and identity online?". Five alternatives were available, "Rejecting the use of cookies" - "Chosen not to use a website due to uncertainty regarding the use of collected data" - "Provided false or fictive information during registration on a website" - "Requested one or more websites to not share information with third-parties" - "Requested one or more websites to delete all personal data stored about me".

4.2.3 Control Group

Background

Interpreting the degree of awareness and risk perceptions for digital natives can be complicated and difficult by solely investigating the digital natives as a group alone. However, the use of a control group is useful to isolate the independent variable's effect and then examine the dependent variables to rule out or confirm explanations. Ideally, the original sample and the control group are identical, except for the independent variable distinguishing the groups. This can help to understand what factors affect the outcomes, leading to significantly increasing the ability to conclude the study. Having a control group also reduces the possibility of making an erroneous conclusion.

Nettskjema, the same tool used in the first questionnaire, was used to create and gather responses. Moreover, the questionnaire was based on the same questions as the previous questionnaire towards the digital natives and did not need any new validation from NSD to assure that the data collected did not violate the respondents' data privacy.

Conducted Method

The control group chosen for comparison were non-digital natives. These are people born before 1980, which was used to define the digital natives in this study. The non-digital natives chosen for the control group were recruited through acquaintances in a social network. For sample control, the questionnaire was open for everyone with access to the link. However, since the control group could not be authenticated in the same way as the digital natives, the questionnaire's link was not shared widely through forums and social media. Instead, it was provided through the network of non-digital acquaintances towards their connections. Moreover, unlike for the digital natives, the ability to choose between age intervals was open in the control groups' questionnaire. This increased the sample control by removing any answers in the wrong age category after the data collection was over. These measures limit the possibility for multiple answers and the opportunity for randoms to influence the results by providing fake answers.

The questionnaire for the control group was created as a separate questionnaire, to prevent mixing and misinterpretation when extracting the results. Further, the control group of non-digital natives was asked the same questions as in the first questionnaire towards the digital natives, except the question towards faculty belonging. Since there was no assurance that the non-digital natives had studied at NTNU, excluding the question prevented confusion and misunderstandings.

4.3 Data Analysis

To analyze and present data, a predefined method was conducted. The analysis followed the same recipe for each method and presented the data in the same structured way. This created a good structure, with a clear step in each analysis part. For the reader, it results in similar patterns, which would help build an easier understanding and follow along through the analysis. Firstly, the analysis will be described with an introduction. Then, the part of the actual analysis will consist of results from a descriptive analysis. Secondly, a bivariate analysis was conducted. The results were combined with the results from the descriptive analysis to strengthen the analysis. Through the analysis, concrete results from the tests were reported, and the results were presented through suited graphs, such as bar graphs, scatter plots, and pie graphs.

4.3.1 Content analysis

Data returns as a measurement of a variable after the data was collected through the established method. The content analysis was presented through a structured and categorized method for comparing the results. Moreover, these results were placed within a taxonomy. The taxonomy will be a slight variation from the taxonomy seen in the paper by Earp, J. et al. [44]. Furthermore, the taxonomy con-

tains a collection of each business model and the specific websites within them.

4.3.2 Questionnaire

Data analysis on the results from the questionnaire was done using the statistical program "SPSS". SPSS allowed the solving of a variety of business and research problems. *It provides a range of techniques including ad-hoc analysis, hypothesis testing and reporting – making it easier to manage data, select and perform analyses, and share your results.* For the questionnaire, the primary two techniques were used to measure variables: checklists and rating scale/Likert scale. Checklists contained descriptions, where the respondent indicate if the item listed were present/true or not present/false, depending on the question. Rating scale, on the other hand, presents answers on a scale e.g., never to always, or in terms of numbers from 1 to 4.

To analyze the data from the questionnaire, frequencies from the questions were first presented; for example, the percentage from each age group or gender distribution from the respondents. Furthermore, frequencies from different groups could be combined to illustrate how many were represented in a given answer on another question. Values, such as percentage, count and mean were used to illustrate the responses given for the questions. To examine differences between groups, ANOVA one-way analysis of variance, was conducted to compare the means of two or more samples. Even though this method of comparing data has received criticisms from scientists with arguments that the method is not suited for ordinal data and smaller samples. ANOVA one-way is utilized in this thesis, based on the conclusion from Geoff Norman's paper [73], which states *Parametric statistics can be used with Likert data, with small sample sizes, with unequal variances, and with non-normal distributions, with no fear of "coming to the wrong conclusion". These findings are consistent with empirical literature dating back nearly 80 years.*

4.4 Ethical and legal considerations

Even though this study gathered information from a variety of sources such as privacy policies from websites, and individuals through the use of questionnaires. No personal identifiable data was stored, neither on local systems or cloud servers connected to external services used for data collection. This was confirmed by NSD, after the thesis was reported with the questionnaire attached for a review. The results from NSD were clear, as this study did not collect any personal identifiable data and were not obligated to inform the users. Either way, the respondents in both questionnaires were given the opportunity to reach out and request the deletion of their response if wanted. This was not necessarily needed, but done to respect the respondents and give the right to change their mind.

Being personally attached to the phenomenon can lead to unfortunate effects on

the results of a study. However, this has not been the case for this study. It was made clear from the start that an objective view must be emphasized to produce results with credibility. Further, the use of sources from the internet requires probity when choosing what and whom to cite. This thesis has a responsibility for its credibility to avoid plagiarism, manipulation, and forging of data. Not only would that violate proper scientific practice, but it also ruins the integrity of the thesis. To avoid this, there has been empathized a good practice though out the thesis.

Chapter 5

Results

In this chapter, the results from the conducted methods will be presented. First, a visual presentation of the results from the content analysis is shown in a taxonomy. Following, a more in-depth descriptive analysis is presented to highlight some of the findings from the content analysis. After the content analysis, the results from the initial questionnaire is presented with visuals such as graphs and tables combined with descriptive texts to present the results. Lastly, the same method as with the initial questionnaire is conducted on the control group, before combining the initial questionnaire with the control group to highlight important results.

5.1 Results from Content analysis

Website	Data Collection			Third-party Data Collection			Data Storage			Data Sharing			Marketing Communication			Total questions	Questions Answered
E-commerce:	*yes*	*no*	*dno*	*yes*	*no*	*dno*	*yes*	*no*	*dno*	*yes*	*no*	*dno*	*yes*	*no*	*dno*	*34*	max(34)
Elkjop.no	10	0	2	4	0	0	6	0	2	6	0	1	1	0	2	34	27
Zalando.no	12	0	0	4	0	0	6	0	2	7	0	0	1	0	2	34	30
Ebay.com	9	0	3	4	0	0	7	0	1	6	0	1	2	0	1	34	28
Komplett.no	11	0	1	4	0	0	7	0	1	7	0	0	1	0	2	34	30
Ikea.no	12	0	0	4	0	0	8	0	0	7	0	0	1	0	2	34	32
Blogs:																	29,4
Blogg.no	12	0	0	4	0	0	8	0	0	6	1	0	1	0	2	34	32
Blogger.com	12	0	0	4	0	0	8	0	0	6	1	0	1	0	2	34	32
Squarespace.com	12	0	0	4	0	0	8	0	0	6	1	0	1	0	2	34	32
Wordpress.com	10	0	2	4	0	0	7	0	1	7	0	0	1	0	2	34	29
Wordpress.org	12	0	0	4	0	0	7	0	1	7	0	0	1	0	2	34	31
News sites:																	31,2
Vg.no	9	0	3	4	0	0	5	0	3	4	0	3	0	0	3	34	22
Dagbladet.no	12	0	0	4	0	0	5	1	2	6	1	0	3	0	0	34	32
Aftenposten.no	9	0	3	4	0	0	5	0	3	4	0	3	0	0	3	34	22
Nettavisen.no	12	0	0	4	0	0	7	1	0	6	1	0	3	0	0	34	34
Dn.no	10	0	2	4	0	0	7	0	1	6	0	1	1	0	2	34	28
Social Media:																	27,6
Facebook.com	12	0	0	4	0	0	7	0	1	6	1	0	1	2	0	34	33
Snapchat.com	8	0	4	4	0	0	6	0	2	5	1	1	1	0	2	34	25
Instagram.com	12	0	0	4	0	0	7	0	1	6	1	0	1	2	0	34	33
Linkedin.com	10	0	2	4	0	0	7	0	1	5	1	1	1	2	0	34	30
Twitter.com	10	0	2	4	0	0	6	0	2	5	1	1	1	0	2	34	27
Sum	216	0	24	80	0	0	134	2	24	118	10	12	23	6	31		29,6
																Average:	29,45

Figure 5.1: Content Analysis Taxonomy

The non-Norwegian websites seemed to have the same amount of questions answered as the Norwegian policies. The average total number of questions answered was 29,55 out of 34, which is 86.91% questions answered, resulting in 13,09% of total questions that could not be answered. Vg.no and Aftenposten had the lowest amount of questions answered with 22 of out 34 in their policy, which can be explained by as they use the same provider for their privacy policies.

Blogs had the most questions answered overall with a mean of 31,2 out of max 34. However, none of the blog policies mentioned anything about if they or third parties could send unwanted emails to the users of their service. This seemed to be the trend throughout all business models, except for social media, where 3/5 policies explained clearly that they or third parties would not send unwanted emails. Which can be explained by Facebook and Instagram use the same policy, but maybe also be because social media companies have been under massive pressure due to previous privacy incidents.

Furthermore, social media and blogs, as well as 2 news sites, were not refraining from selling consumers data. Every single social media policy stated that they did sell consumers data to third parties, while none of the e-commerce websites did. It mostly consisted of questions towards the protection of the data stored, both offline and against unauthorized access. News sites were the only business model with occasions of policies stating that they would keep data about users stored even after the user deleted the account. Furthermore, policies in the news site business model also tend not to be able to answer fully in the Data storage category.

Overall, all policies mentioned third-parties and their use of them. Resulting in all 20 policies answering “yes” on the 4 questions in the Third-party data collection category. All websites collected PII about their users, as well as aggregate information. Most of the websites also collected geolocation and data from inquiries. Moreover, all websites share this information (both PII and aggregate data) within the company, as well as with third-party companies. Almost all websites use personalized advertisement and can send their consumers newsletters with their consent (vg.no only policy not stating anything about newsletters).

Regarding the GDPR, every single website policy included and mentioned the users right regarding the law. Indicating that the policies within all business models are up to date and fulfills the requirement to the GDPR.

5.2 Results from questionnaires

This section contains the results from the initial questionnaire towards the digital natives. In total the questionnaire had a number of 96 respondents, all within the target group. Essentially, the total number of answers for each question was N=96, otherwise it would be remarked for the particular question.

5.2.1 Demographics

Age

Since the target group was digital natives, which consists of people born after 1980, all of the respondents were within this age group. The large majority of the respondents with 83,33% were 19-29 years old, while the rest of the respondents 16,67% were 30-39 years old.

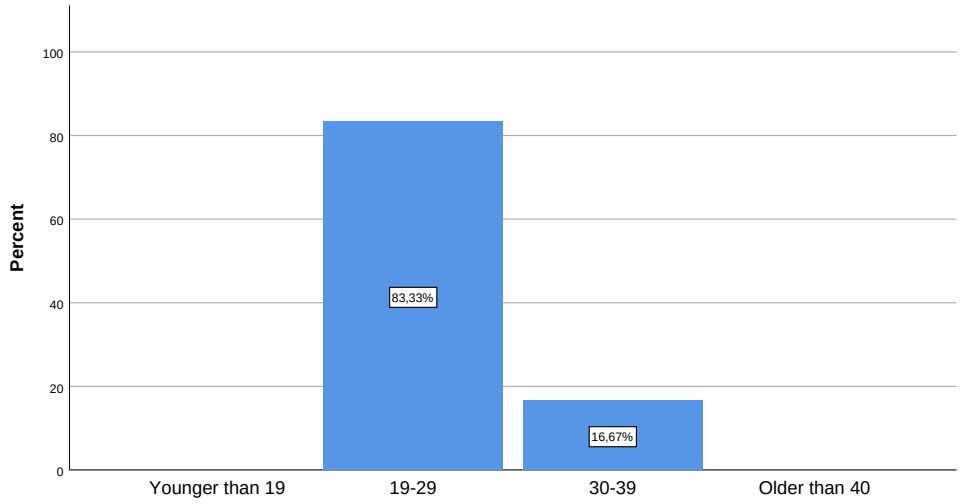


Figure 5.2: Age distributions in % for the questionnaire respondents.

Gender

The distribution of gender in the questionnaire was 41,7% women and 58,3% men. Showing that there was a majority of male respondents with 13,4% more than women from the students in this questionnaire.

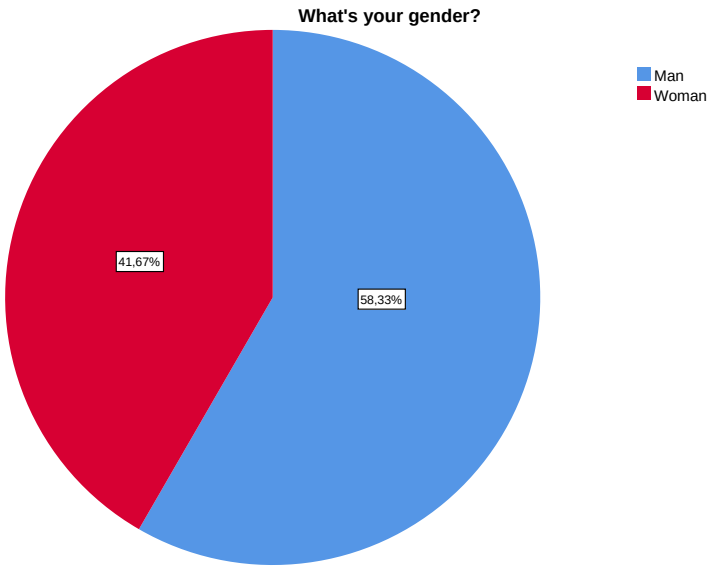


Figure 5.3: Gender distributions in % for the questionnaire respondents.

5.2.2 Background information

Education

There was a balanced sample from each of the groups of "education level achieved" by the respondents. Out of the total answers, 34,4% had finished "videregående skole", while 33,3% had finished a higher education up to 4 years. Further, 29,2% (n=28) answered with a degree of education of more than 4 years study. Lastly, only 1,04% answered "fagskolenivå", while 2,08% commented with "other" levels of education.

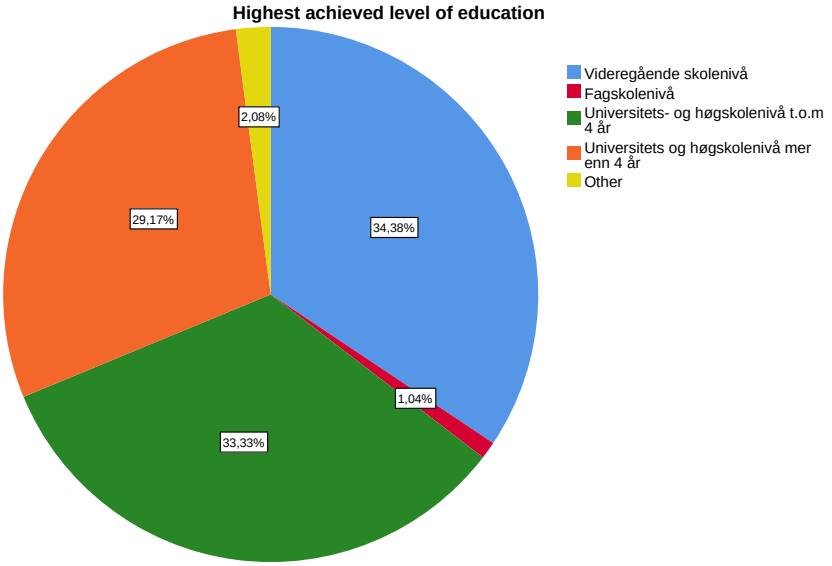


Figure 5.4: Highest achieved education distributions in % for the questionnaire respondents.

Information Security Experience

Looking at the information security experience the student respondents had, there were some experience within the different categories. This question was a phrased in a multiple-choice fashion, letting the respondents choose more than one category if they have experience from several sources. However, one of the more interesting findings is the amount of students with "No experience". The overall experience in information security seems to be either no experience or a hobby/interest. 34,56% of the total respondents, answered that they have no information security experience.

	Course	Work Related	Education	Subject	Hobby/Interest	No Experience	Other
	14	0	0	0	0	0	0
	0	26	0	0	0	0	0
	0	0	19	0	0	0	0
	0	0	0	17	0	0	0
	0	0	0	0	32	0	0
	0	0	0	0	0	36	0
	0	0	0	0	0	0	5

Figure 5.5: Information security experience distribution for the questionnaire respondents.

By further examining the students with no information security experience, the respondents without information security experience was linked to their highest

achieved level of education. As seen in figure 5.6 below, of the total answers, "Videregående" students had the least experience, while digital natives with "a degree of education of more than 4 years study" had less, and lastly the most educated digital natives with "a degree that took more than 4 years of higher education" had the least respondents with no information security experience.

		No Experience
Highest achieved level of education	Videregående skolenivå	16
	Universitets- og høgskolenivå t.o.m 4 år	12
	Universitets og høgskolenivå mer enn 4 år	8

Figure 5.6: Highest education level and No Experience for the questionnaire respondents.

Browsing habits

The respondents were presented with a question related to their time spent online browsing the internet in order to examine the browsing habits of the digital natives. The respondents had to estimate how many hours they spent online every day by choosing an answer within an interval ranging from "0" to "7 or more". The results can be seen in figure 5.7, and showed that the majority of the respondents with 50% spent 3-4 hours online every day. There is a deceasing amount of people with more hours online, 25% answered with 5-6 hours every day, while 15,6% said they spent 7 or more hours online. On the other end of the scale, only 9,4% spent 1-2 hours every day, and no one is offline every day.

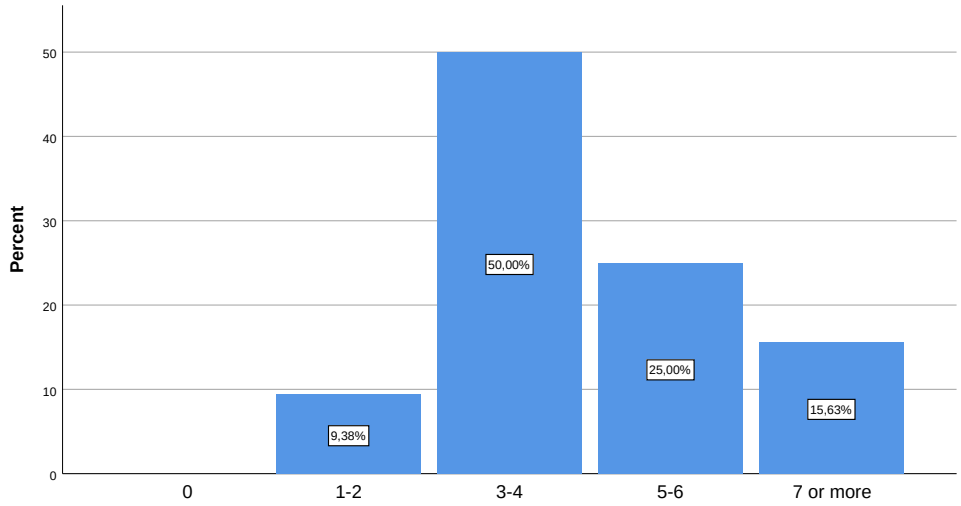


Figure 5.7: Estimated hours online every day for the questionnaire respondents.

5.2.3 Security Awareness

Awareness about information gathering

To measure awareness regarding information gathering, the respondents were asked questions about terms related to data privacy. As mentioned in the methodology chapter, a definition of the term was included in the description of the question for validation of the awareness of the respondents.

The first question: "Do you know what a Cookie is?", received a large majority with 76,04% answers of "Yes, I am fully aware". Furthermore, 20,83% answered "Yes, I am partly aware", while only 1,04% "only have heard about it" and 2,08% was "not aware" of what a cookie is. Between the genders, men had a mean value of 1,16 versus women's mean value of 1,48, this was not enough to prove a significance difference between the genders. Furthermore, the independent-samples T Test shows no significant effect for gender, $t(94) = -2.621$, $p=.010$, despite Men ($M = 1.16$, $SD = .417$) attaining lower scores than women ($M = 1.48$, $SD = .751$). Moreover, there was neither a significant effect on education level and the awareness of cookies at $p<.05$ for [$F(2,90) = .692$, $p = .503$].

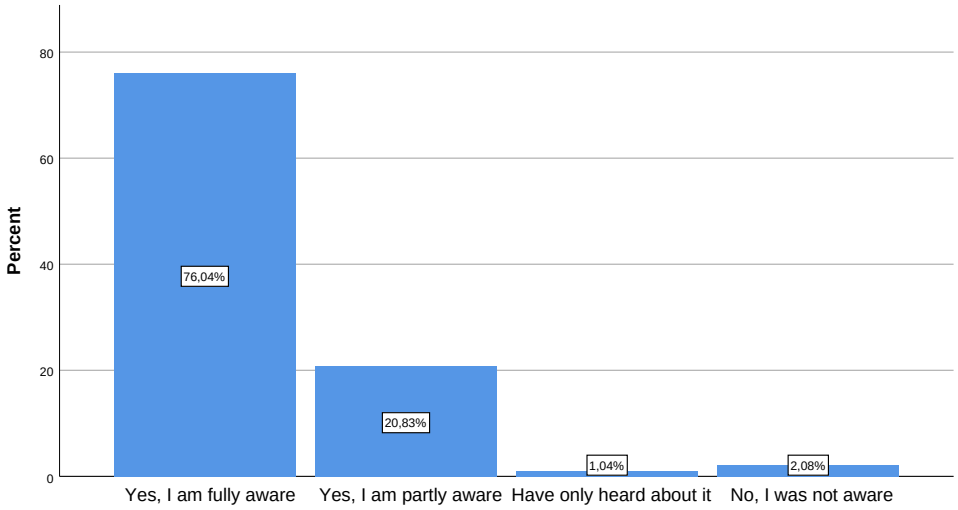


Figure 5.8: Cookie awareness of the questionnaire respondents. (N=96)

The second question: "Do you know what a data broker is?", as seen in figure 5.9 received a spread response with 32,29% answers of "Yes, I am fully aware". Furthermore, 29,17% answered "Yes, I am partly aware", while 15,63% "only have heard about it" and 22,92% was "not aware" of what a data broker is. With an ANOVA one-way test, there was found a significant effect on faculty and the awareness of data brokers at $p<.01$ for $F(7, 80) = 4.267$, $p = 0.000$. Also, by conducting the same Anova one-way test, this resulted in a significant effect on amount of hours spent online and awareness of data brokers at $p<.05$ for $F(3, 92) = 4.805$, 0.04. The results are illustrated in figure 5.10 A deeper look at those variables with a bivariate correlation, shows there was a negative correlation between the amount of hours spent online every day and awareness of data brokers, $r = -.348$, $n = 96$, $p = .001$. A scatter-plot summarizes the result in figure 5.11 Lastly, there was no correlation between the education level and the awareness of data brokers $r = -.207$, $n = 93$, $p = .132$.

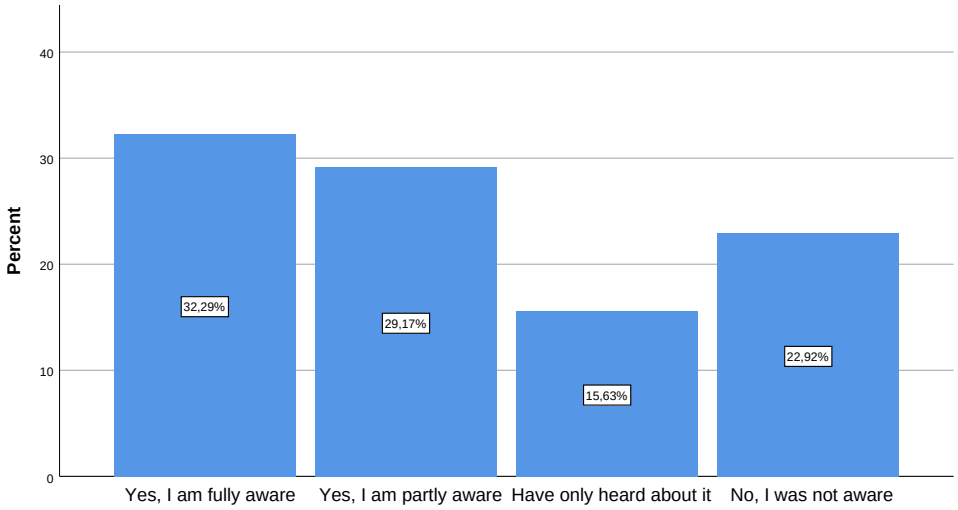


Figure 5.9: Data broker awareness of the questionnaire respondents.

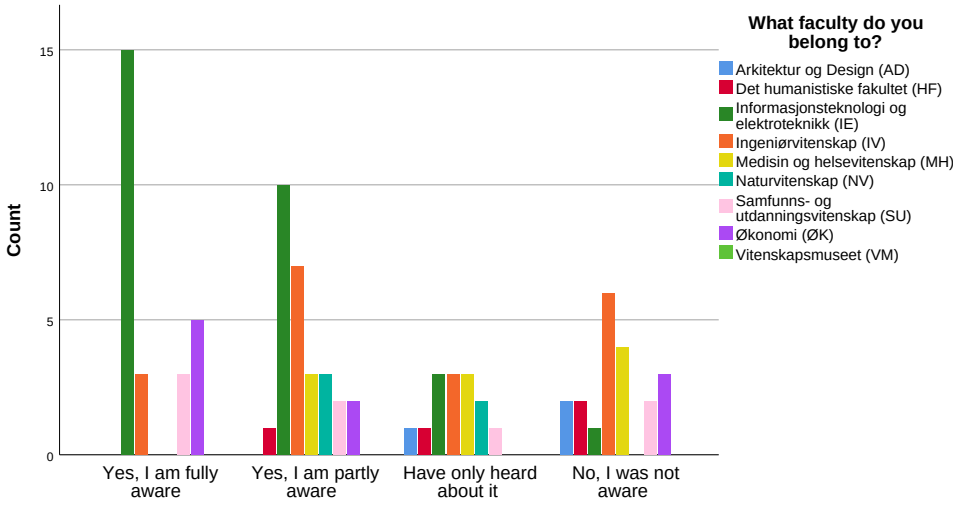


Figure 5.10: Data broker awareness and faculty of the questionnaire respondents.

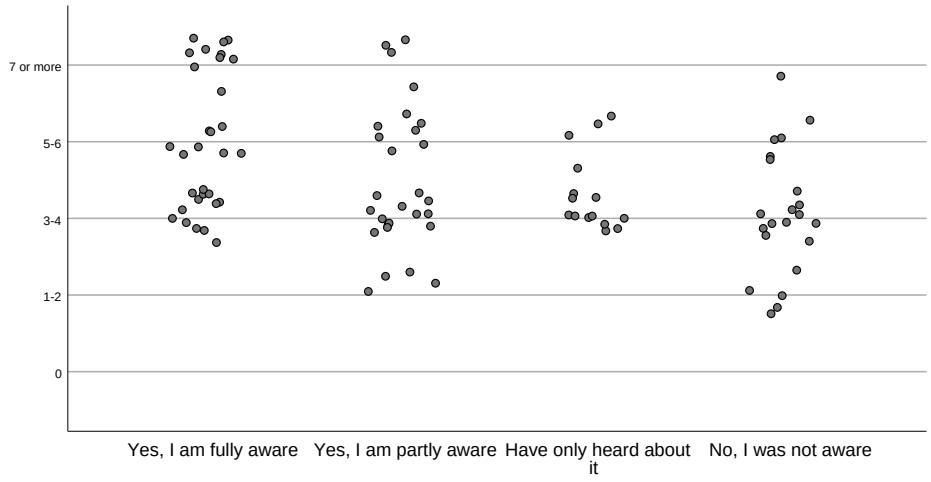


Figure 5.11: Data broker awareness and amount of hours spent online every day of the questionnaire respondents.

The question: "Do you know what a web beacon is?", received 20.83% answers of "Yes, I am fully aware". Furthermore, 18.75% answered "Yes, I am partly aware", while 18,75% "only have heard about" and the majority with 41.7% was not aware of what a web beacon is. Between the genders, men had a mean value of 2,52 versus women's mean value of 3,23. Using ANOVA one-way this was enough to prove a significant difference between the genders and their awareness about web beacons $F(1,94) = 8.920$, $p = .027$. This was also the conclusion of the independent-samples T Test which showed a significant effect for gender, $t(94) = -2.987$, $p = .004$, with Men ($M = 2{,}52$, $SD = 1.206$) attaining lower scores than women ($M = 3.23$, $SD = 1.050$). Moreover, there was a not significant effect on higher education and the awareness of cookies at $p < .05$ for $F(2,90) = 1.144$, $p = .323$. Lastly, there was no correlation between the amount of hours spent online every day and the awareness of web beacon $r = -.239$, $n = 93$, $p = .019$.

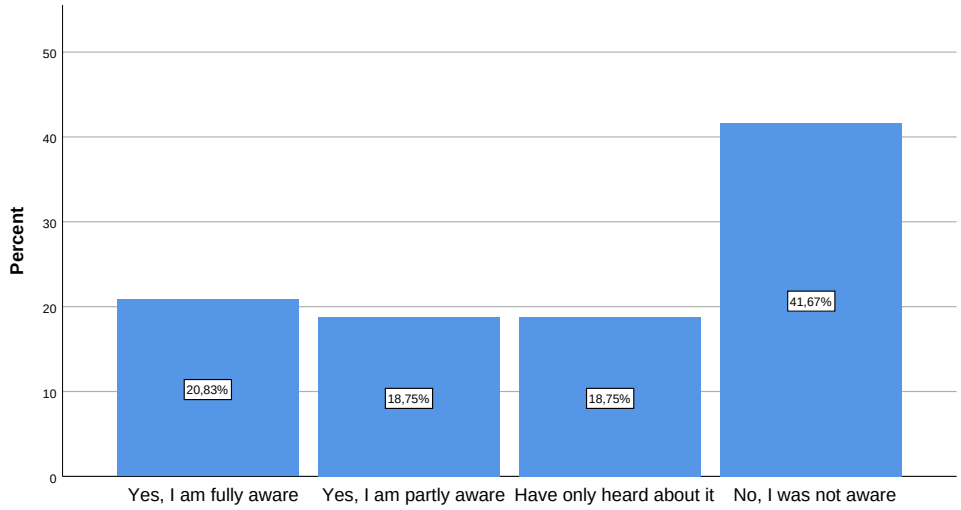


Figure 5.12: Web beacon awareness of the questionnaire respondents.

Lastly, the figure 5.13 below, shows how the mean value for decreases for the awareness of the terms "cookie" and "web beacon" when the level of education increase. However, on the term "data broker", the mean value decrease from "videregående" to until "degree for less than 4 years of educational" and then the mean value increase by 0.15 from "degree for 4 years of educational" to "degree for more than 4 years of education".

		N	Mean
Do you know what a cookie is?	Videregående skolenivå	33	1,39
	Universitets- og høgskolenivå t.o.m 4 år	32	1,28
	Universitets og høgskolenivå mer enn 4 år	28	1,21
	Total	93	1,30
Do you know what a data broker is?	Videregående skolenivå	33	2,70
	Universitets- og høgskolenivå t.o.m 4 år	32	2,06
	Universitets og høgskolenivå mer enn 4 år	28	2,21
	Total	93	2,33
Do you know what a web beacon is?	Videregående skolenivå	33	3,09
	Universitets- og høgskolenivå t.o.m 4 år	32	2,81
	Universitets og høgskolenivå mer enn 4 år	28	2,64
	Total	93	2,86

Figure 5.13: Mean values for the awareness of the questionnaire respondents.

Awareness about user's rights online

To examine if the respondents were aware of their rights when browsing the internet, a general question related to GDPR was conducted. The respondents were presented with a summary of their rights in a bullet list, to validate their own awareness. The question: "Are you aware of your rights in relation to GDPR?", received a majority with 45,26% answers of "Yes, I am fully aware". Furthermore, 27,37% answered "Yes, I am partly aware", while 10,53% "only have heard about" and the majority with 16,84% was not aware of what their rights are. There is not a significant difference between level of education and awareness about GDPR rights, the ANOVA one-way test gave $F(2, 90) = 1.144$, $p = .145$.

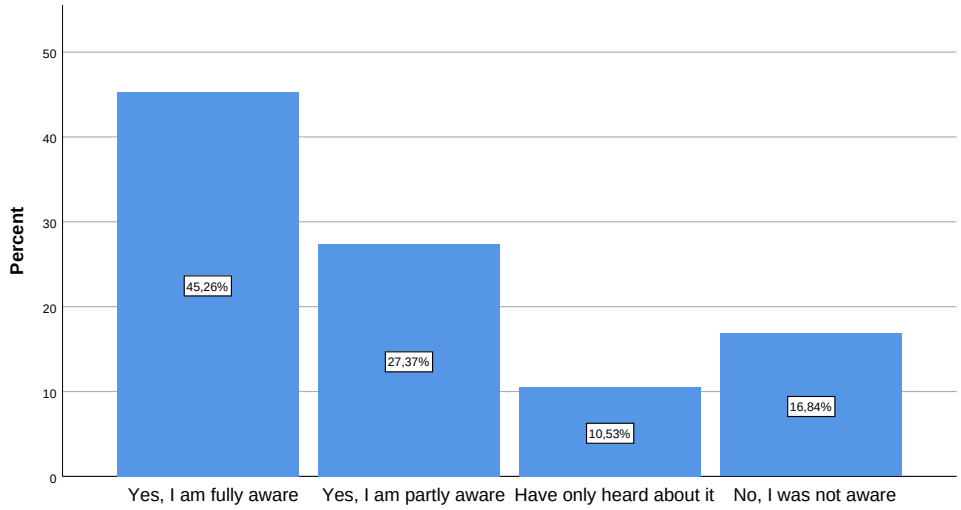


Figure 5.14: GDPR rights awareness of the questionnaire respondents.

5.2.4 Risk perception and willingness

Last part of the questionnaire was used to measure the digital native's risk perception and willingness to accept risk for access to different kinds of websites. The questions are divided into two sub-chapters: "before accessing websites" and "Willingness to accept risk". Each sub-chapter, contains results from the questionnaire for the questions within the sub-chapters.

Before accessing websites

The respondents were presented with a picture of a cookie pop-up, with the regular options users get when accessing a website. Figure 5.15 shows the distribution of the answers made by the users, where a large majority with 62,5% accepts all cookies. On the other hand, 3,13% only removes marketing cookies and pixels, while 34,38% only accept the use of strictly necessary cookies. None of the respondents only removed personlization cookies. Furthermore, there was not a significant difference between level of education and cookie acceptance, the ANOVA one-way test gave $F(2, 90) = .857$, $p = .428$. Neither was it a significant difference between amount of hours online every day and cookie acceptance, ANOVA one-way test gave $F(3,92) = .561$, $p = 6.42$.

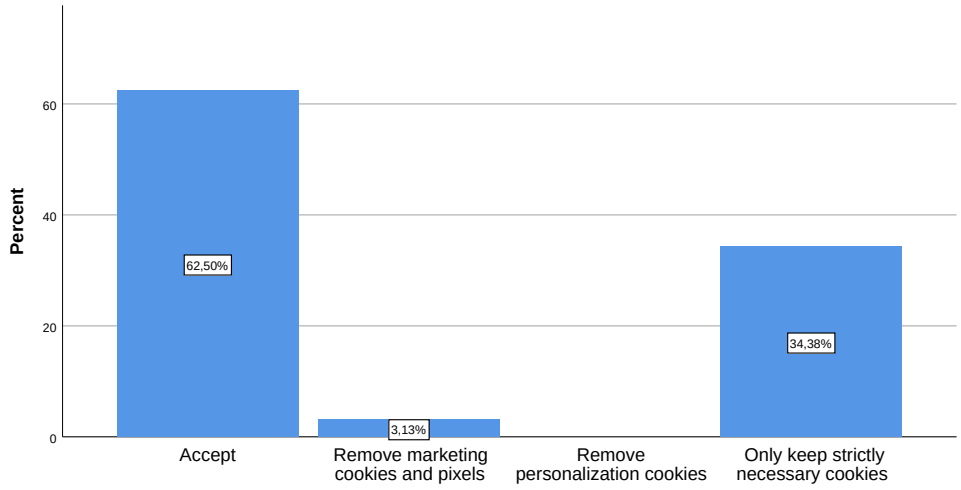


Figure 5.15: Cookie acceptance of the questionnaire respondents.

To see if the respondents read the policy before they accessed websites, they were asked the question: "Do you read the privacy policy before accessing a website?". The results showed that only 3,13% always read the policy, and 38,54% sometimes read it. Lastly, the majority with 58,33% never read the policy before accessing a website. The distribution can be seen in figure 5.16. Further, there is not a significant difference between "level of education" and "reading the policies before accessing websites", the ANOVA one-way test gave $F(2, 90) = 5.116$, $p = .008$. While Post hoc comparisons using the Tukey HSD test indicated that the mean difference for the "videregående" education level and higher education with (Mean difference $= .350$, $p = .27$) to "university education less than 4 years" and (Mean difference $= .395$, $p = .015$) for "university education longer than 4 years". Neither is there a significant difference between amount of hours online every day and reading policies before accessing websites, ANOVA one-ways test shows: $F(3, 92) = 1.357$, $p = .261$.

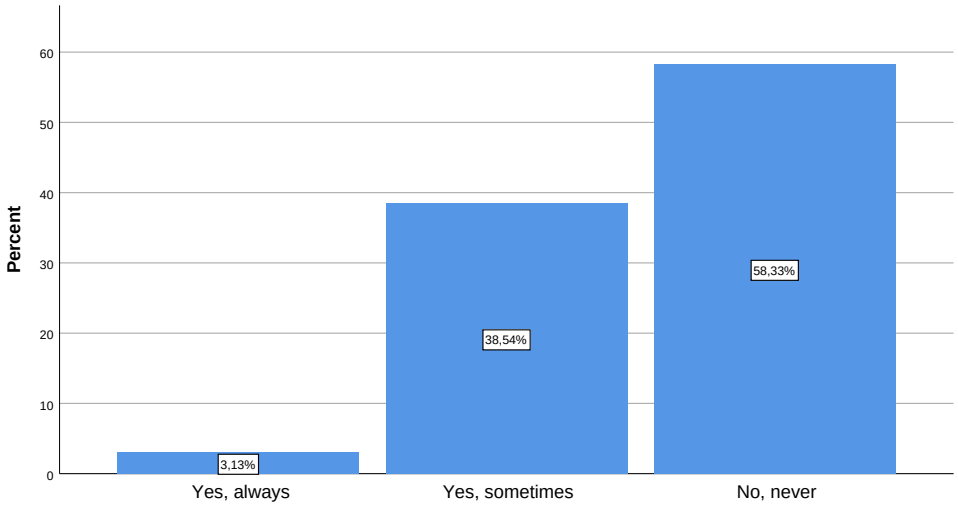


Figure 5.16: Reads policy of the questionnaire respondents.

Willingness to accept risk

The willingness to accept risk and provide information were measured by scenarios based on the content analysis. The questionnaire included 6 different scenarios where the respondents were provided with a risk connected to a benefit with accessing a service or website. The answers were represented with a scale of willingness with values 1-4 from very unwilling - unwilling - willing - very willing.

In the first scenario regarding the respondents willingness to register a user on social network site, with a benefit of access to a network to share pictures, comments and keep in touch with friends and acquaintances. The risk associated was that the social network site had agreements with third party companies for direct marketing towards your preferences and actions while using the site. Only 3,13% of the respondents were "very unwilling", while 18,75% was "unwilling". On the other hand, a large majority with 55,21% was "willing", and 22,92% was "very willing". There was not a significant difference between level of education and scenario 1, ANOVA one-way showed: $F(2,90) = .688$, $p = .505$. There was also no correlation of amount hours spent online every day and scenario 1, $r = -.017$, $n = 96$, $p = .867$.

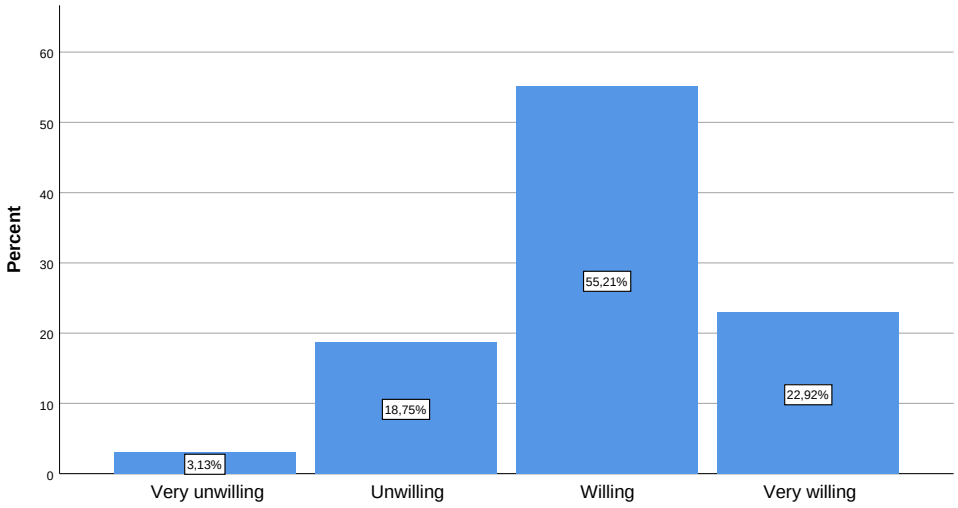


Figure 5.17: Scenario 1: The willingness to register a user on a social network site of the questionnaire respondents.

In the second scenario regarding the respondents willingness to create an account on an online newspaper, with a benefit of access to articles behind a paywall, access to the news archive and access to electronic version of the paper. The risk associated was that the company keeps data stored about users even after the account was deleted, where the PII about users were kept and sent to business partners for profiling. The results shows that only 6.25% were "very unwilling", while 37,5% were "unwilling". The majority with 44,79% were "willing", and 11,46% responded with "very willing". There was no significant difference between level of education and scenario 2, ANOVA one-way shows: $F(2,90) = .749$, $p = .476$. Neither was there a significant difference with faculty belonging, ANOVA one-way: $F(7,80) = .996$, $p = .440$. No correlation was found of amount of hours spent online every day and scenario 2, $r = .021$, $n = 96$, $p = .839$.

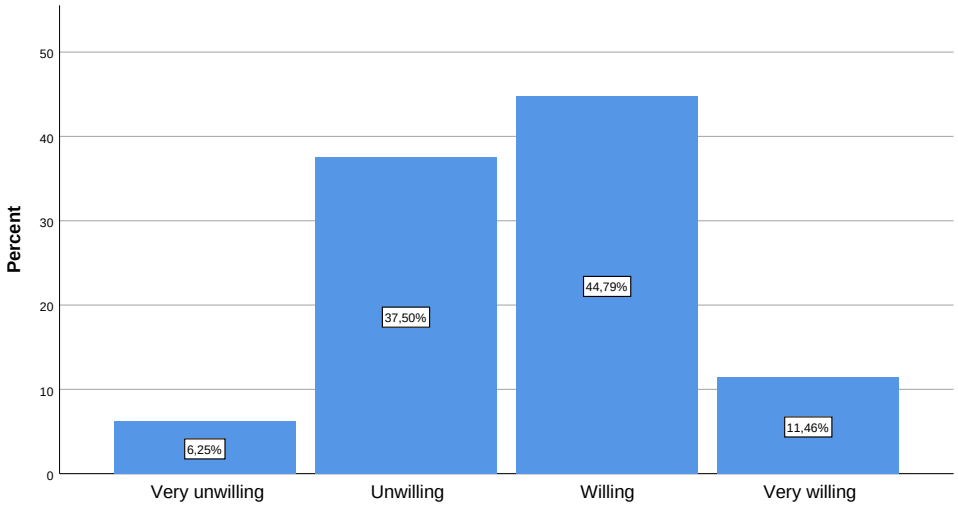


Figure 5.18: Scenario 2: The willingness to create account on online newspapers of the questionnaire respondents.

In the third scenario regarding the respondents willingness to participate in a give-away or survey on blogs, with a benefit of winning prices and support the owner of the blog. The risk associated was that the blog provides users' e-mail to third parties which could lead to unwanted spam from unknown sources. Here the large majority with 50% were "very unwilling", while 37% were "unwilling". On the other hand, 11,46% were "willing", and only 1,04% "very willing". There was no significant difference between level of education and scenario 3, ANOVA one-way gave: $F(2, 90) = .042$, $p = .959$. No correlation was found of amount of hours spent online every day and scenario 3, $r = 0.57$, $n = 96$, $p = .583$.

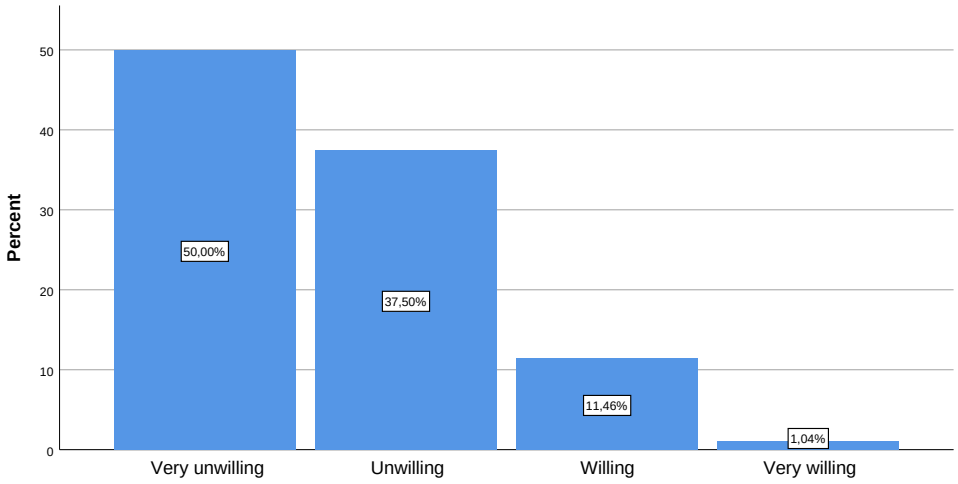


Figure 5.19: Scenario 3: The willingness to give information to blogs of the questionnaire respondents.

In the fourth scenario regarding the respondents willingness to do online shopping, with a benefit of simple access to large selection of goods and the opportunity to shop from home. The risk associated was that the online store saves and provides your preferences to third parties used to send you advertisement for similar goods and sales. Only 3,16% were "very unwilling", and 9,47% "unwilling". The large majority with 48,42% were "willing" and 38,95% "very willing". There was no significant difference between level of education and scenario 4, ANOVA one-way showed: $F(2, 89) = 1.782$, $p = .174$. Between the faculties there was some difference, but not significant with ANOVA one-way: $F(7,79) = 3.117$, $p= .006$.

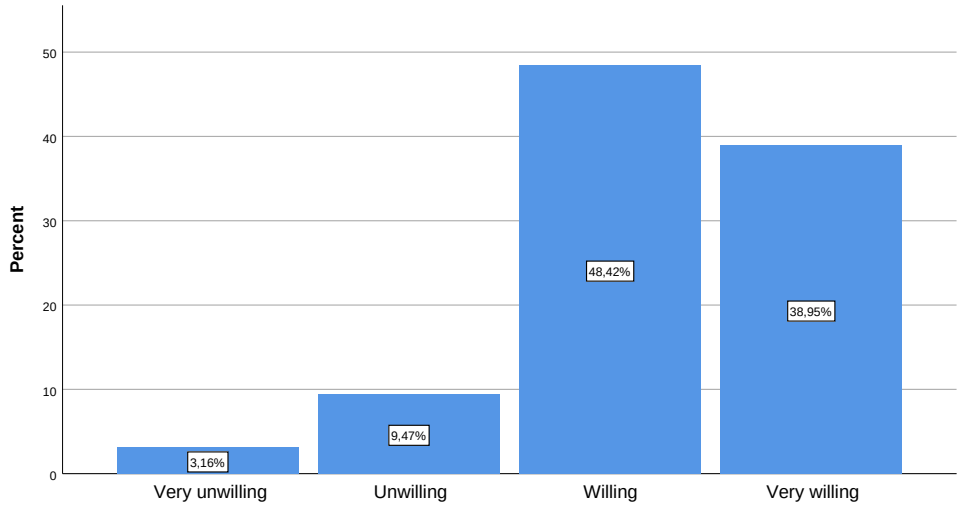


Figure 5.20: Scenario 4: The willingness to do online shopping of the questionnaire respondents.

In the fifth scenario regarding the respondents willingness to use a service that use geolocation, with a benefit of access to easy pinpointing of position without searching, estimated time of arrival, view traffic jams, etc. The risk associated was that the service gets access and stores users' movement, locations and most visited places, and an exact overview of users' location as long as the service is active. Only 3,13% were "very unwilling", while 22,92% were "unwilling". The large majority with 48,96% were "willing", and 25% "very willing". There was no significant difference between level of education and scenario 4, ANOVA one-way shows: $F(2, 89) = 1.782$, $p = .174$. Between the genders there was no significant difference, an indepentent-samples T test gives $T(94) = -2.645$, $p = .010$, with Men (M=2,79, SD=.110) attaining lower scores than women (M=3,20, SD=.103).

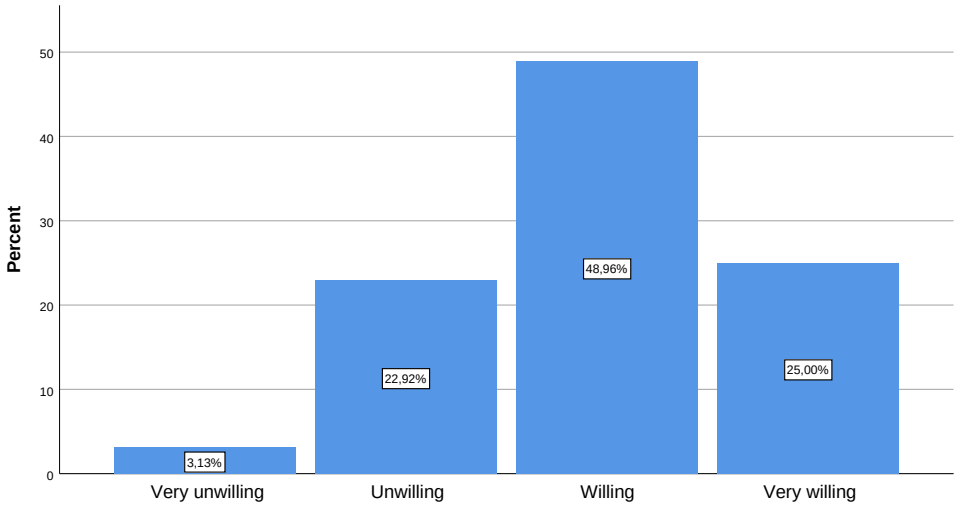


Figure 5.21: Scenario 5: The willingness to use services that track user's geolocation of the questionnaire respondents.

In the last scenario regarding the respondents willingness to participate in a debate on a online forum, with a benefit of being able to debate and discuss with other users, provide information, seek help and help other users. The risk associated was that the forum owns the data of users' posts and sells this data to third parties where statements, attitude, and interests can be used for profiling of the users. There was found no significant difference between level of education and scenario 4, an ANOVA one-way gave: $F(2, 90) = .064$, $p = .938$. However, there was a significant difference between amount of hours spent online every day and scenario 6, ANOVA one-way: $F(3, 92) = 4.806$, $p = .004$. Also, giving a positive correlation between amount of hours spent online every day and scenario 6: $r = .283$, $n = 96$, $p = .005$. Between the genders there was a significant difference, an indepentent-samples T test gives $T(94) = 3.665$, $p < .001$, with Men ($M = 2{,}11$, $SD = .119$) attaining higher scores than women ($M = 1{,}53$, $SD = .088$).

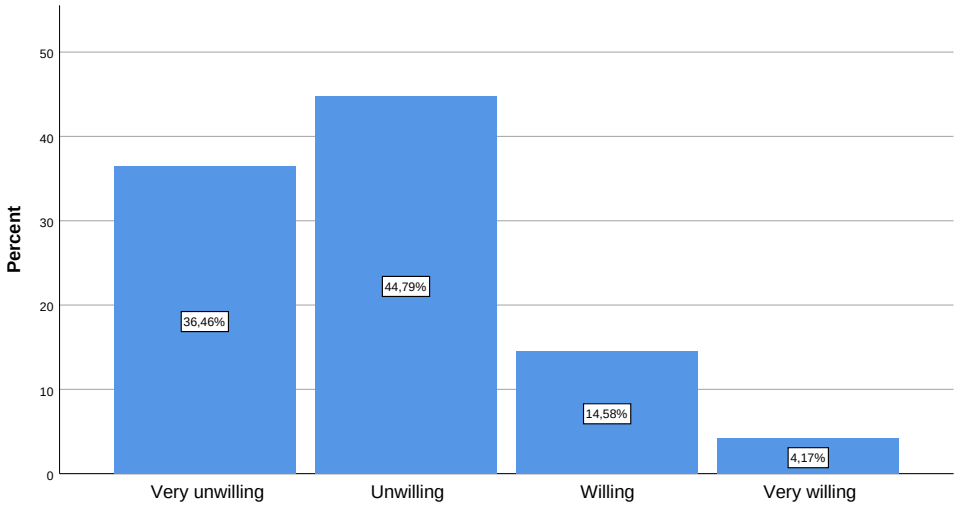


Figure 5.22: Scenario 6: The willingness to debate on a online forum of the questionnaire respondents.

5.3 Results from Control Group

For the control group, the same questionnaire was conducted on another sample as described in the methodology chapter 4.2.3. Essentially, the total number of answers for each question was N=40, otherwise it would be remarked for the particular question.

5.3.1 Demographics

Age

In order to not classify as a digital native, the non digital natives were born before 1980, figure 5.23 show the distribution of the respondents of the control group. Out of the total answers, 17,5% were 40-49 years old, while the majority of the respondents with 45% being 50-59 years old. Lastly, with 37,5% were respondents between 60-69 years old.

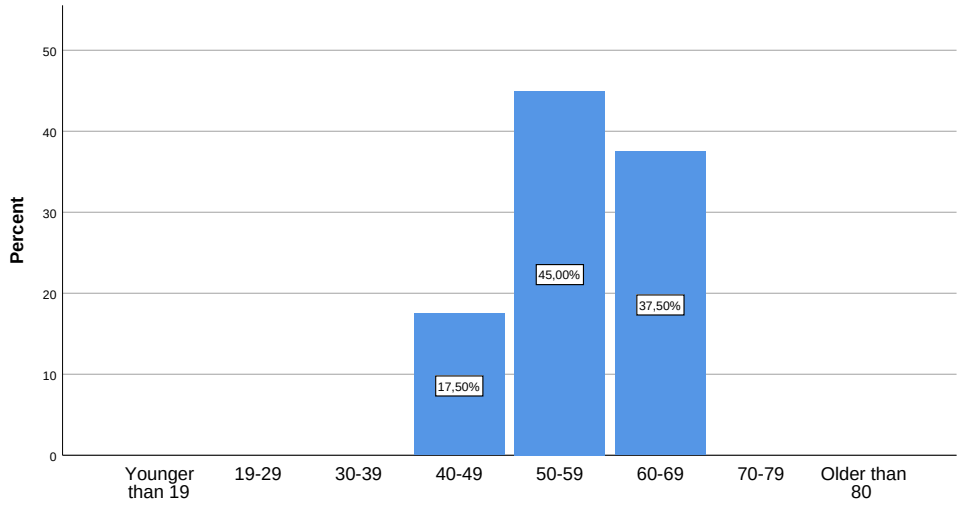


Figure 5.23: Age distributions in % of the control group respondents.

Gender

The gender distribution in the control group consisted of a majority with 77,5% women, and 22,5% men.

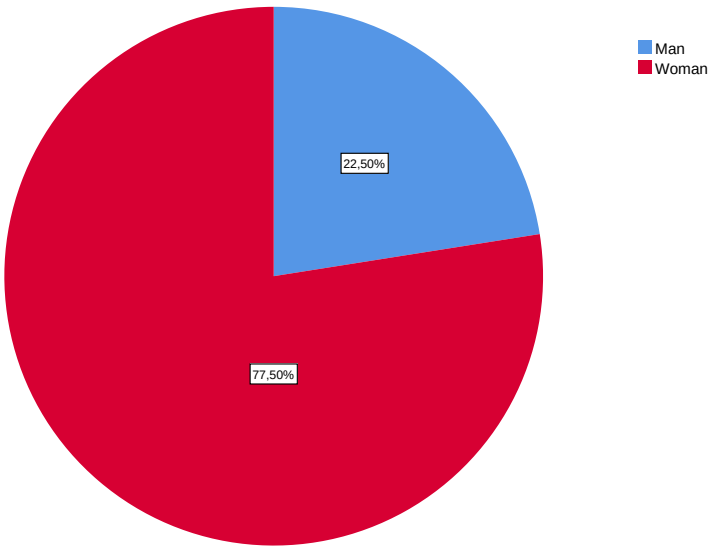


Figure 5.24: Gender distribution in % of the control group respondents.

5.3.2 Background information

Education

Out of total answers, 2,5% had no higher education, 15% finished "videregående skole", and 10% answered "fagskolenivå". While 22,5% had finished a "higher education less than 4 years". Lastly, the majority with 50% answered with a "degree of education of more than 4 years", while no one responded with "other" levels of education.

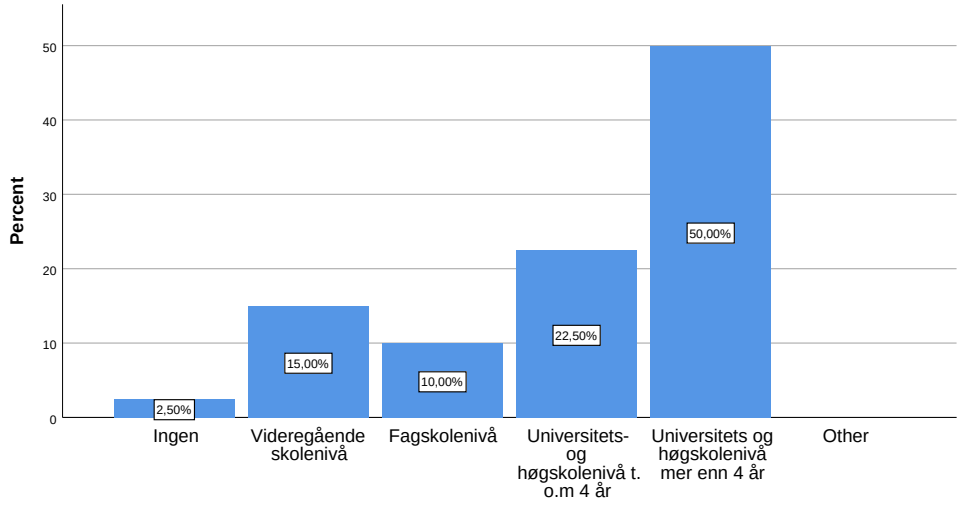


Figure 5.25: Education distribution count of the control group respondents.

By looking at figure 5.26 a comparison is shown by a bar diagram to illustrate the differences in "highest level of education achieved" by both the digital natives and the non digital natives.

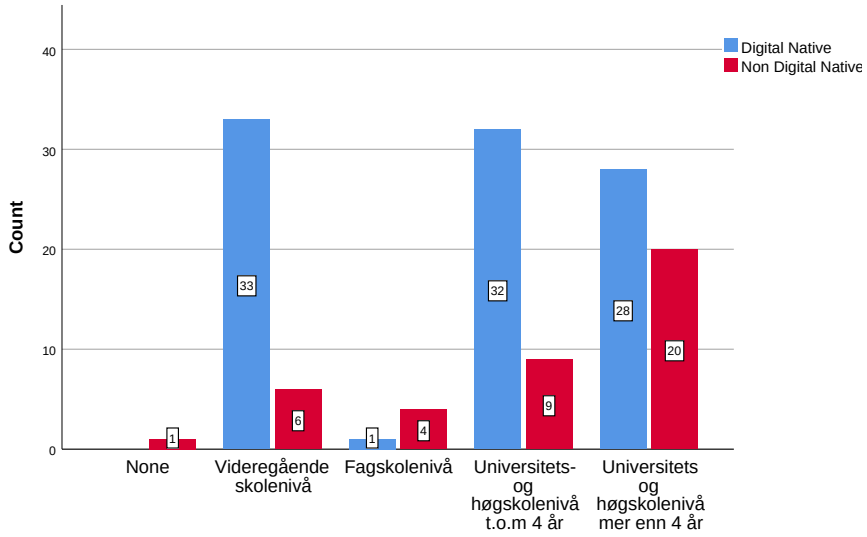


Figure 5.26: Education distribution count of all the respondents.

Information Security Experience

For the control group the information security the majority relates to being work related with 48,72% of the cases. Further, 17,95% of the experience was from a course, 7,692% from education and 7,692% from hobby/interest, while 2,5% was from a subject. On the other hand, 38,46% of the cases had no information security experience at all.

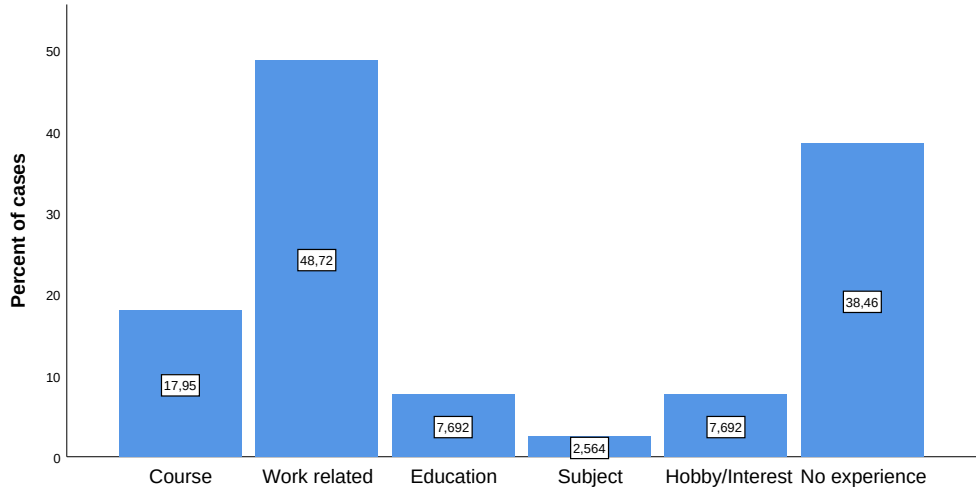


Figure 5.27: Information security experience distribution in % of the control group respondents.

By further examining the non digital natives with no information security experience, the respondents without information security were linked to their highest achieved level of education. As seen in figure 5.28 below, of the total answers, the non-digital natives with more than 4 years of education have the least experience , the non-digital natives with a "degree of education of less than 4 years" had a few less. Following is the non-digital natives with "fagskolenivå", then the respondents with "videregående" as highest achieved level of education. Lastly, no higher education had one case of no experience.

		No Experience
Highest achieved level of education	Ingen	1
	Videregående skolenivå	1
	Fagskolenivå	3
	Universitets- og høgskolenivå t.o.m 4 år	4
	Universitets og høgskolenivå mer enn 4 år	6
	Annet (kommenter under)	0

Figure 5.28: Highest achieved level of education and no information security experience of the control group respondents.

Browsing Habits

For the non-digital natives in the control group, the vast majority with 82,55% spent 1-2 hours browsing every day, while 10% spent 3-4 hours and 7,5% claims to spend 0 hours browsing every day.

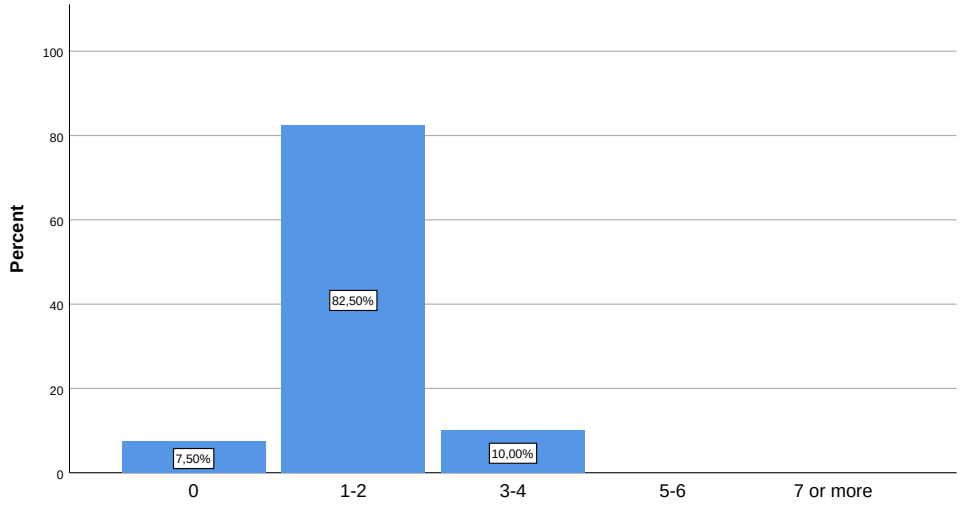


Figure 5.29: Estimated hours online every day for the control group respondents.

ANOVA one-way showed: $F(1, 134) = 99.979$, $p < .001$, showing that there was a significant difference between the digital natives and the non-digital natives regarding how many hours is spent online browsing every day. The scatter plot

below in figure 5.30 illustrates the differences.

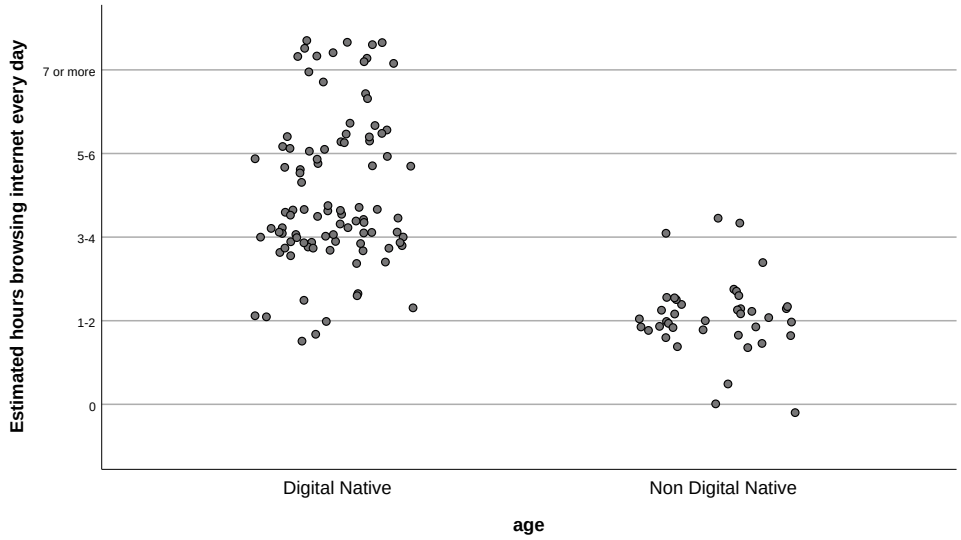


Figure 5.30: Estimated hours browsing every day and age scatter plot for all respondents.

5.3.3 Security Awareness

The control group with non digital natives were asked the about the same terms as the digital natives in the original questionnaire. In the following chapter, a comparison of the digital native's and the non-digital native's answers will be presented and illustrated through charts. Lastly, a table with the ANOVA one-way test results are presented for a overview of the difference between the two groups.

For cookie acceptance, the non-digital natives mostly accept cookies with 68,09% of the respondents. Further, some remove specific cookies where 6,38% removes marketing cookies and pixels, while 6,38% remove personalization cookies. Lastly, 19,15% are only keeping the strictly necessary cookies.

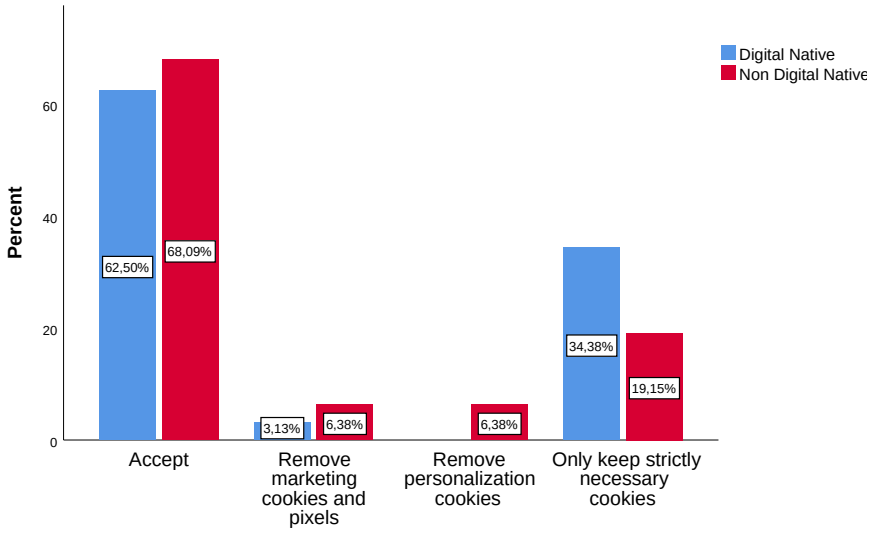


Figure 5.31: Cookie actions of all respondents.

The non-digital natives had a majority of respondents with 58,7% that sometimes read the privacy policy before accessing a website. Furthermore, 6,52% always reads the privacy policy, while 34,78% never reads the privacy policy.

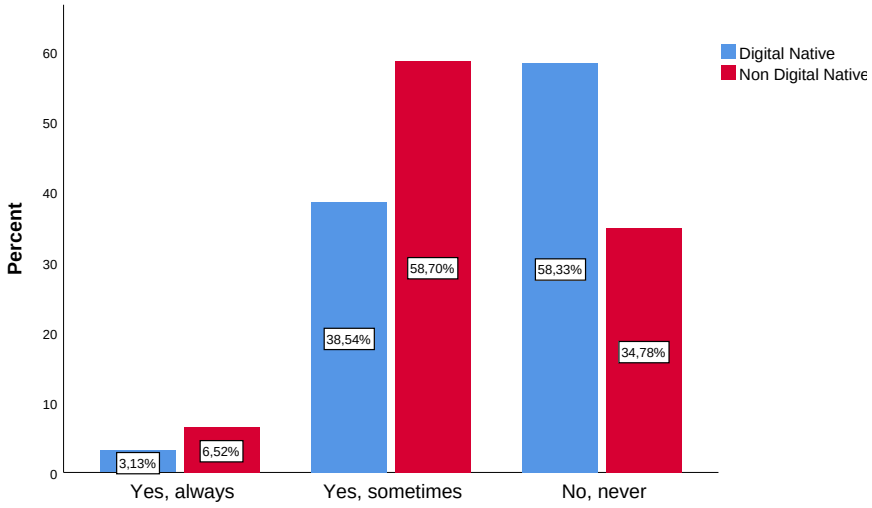


Figure 5.32: Reads the policy of all respondents.

The non-digital natives had a majority of respondents that are "fully aware"

with 55,32%, and 36,17% that were "partly aware". A small amount with 6,38% had "only heard about" cookies, and lastly only 2,13% was "not aware" of cookies at all.

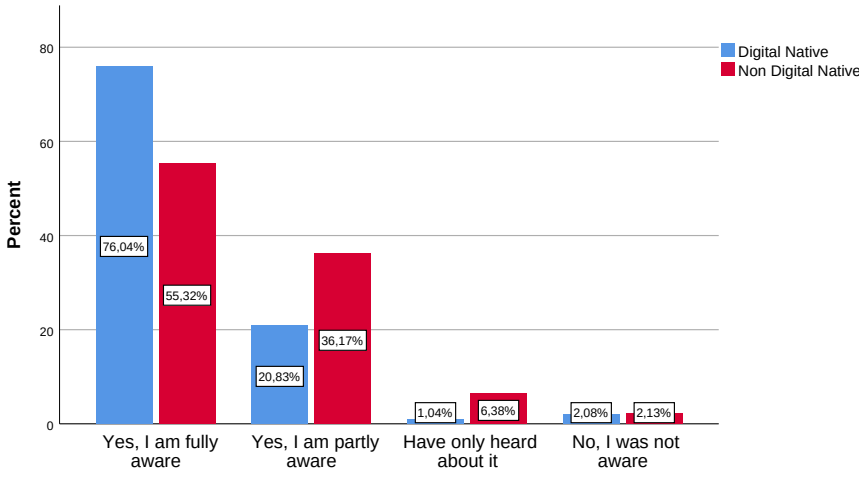


Figure 5.33: Cookie awareness of all respondents.

For awareness about the term data broker, the non digital natives had a small majority with 34,04% that was "not aware". Meanwhile, 27,66% were "partly aware" and 21,28% "fully aware". 17,02% have "only heard about it", which made the response spread from within the group of non-digital natives.

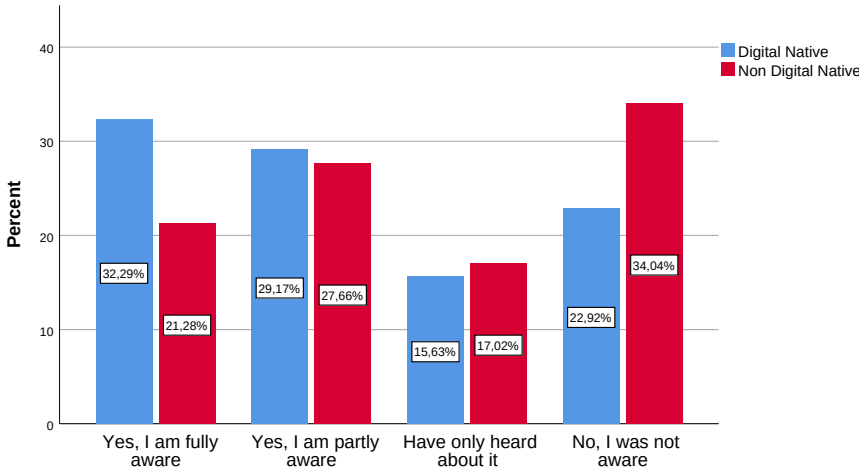


Figure 5.34: Data broker awareness of all respondents.

From the non-digital natives in the questionnaire, only 8,51% were "fully aware" about the term web beacon. 17,02% were "partly aware", while 21,28% had "only heard about it". Lastly, the majority was "not aware" with 53,19% respondents.

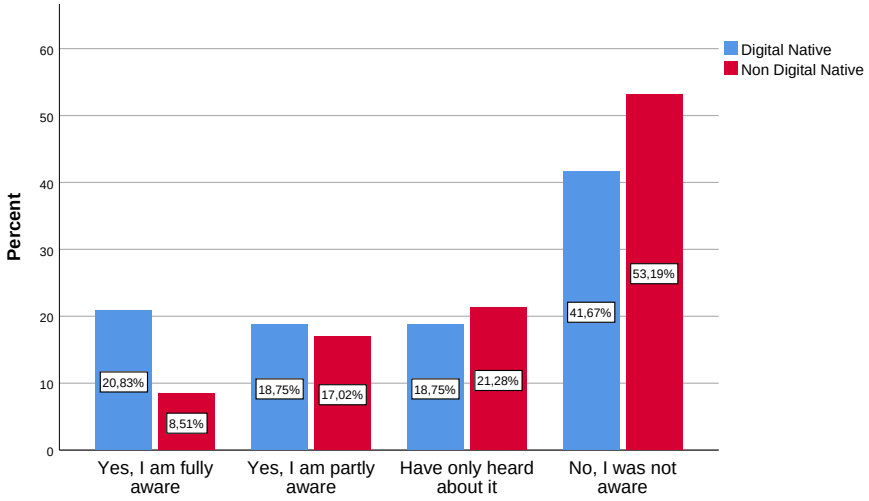


Figure 5.35: Web beacon awareness of all respondents.

For the awareness about user's right online, the questionnaire received a majority of answers from the non-digital natives consisting of "yes I am fully aware"

with 36,17%. Further 27,66% were "partly aware", while 14,89% had "only heard about it". Lastly, 21,28% was "not aware" of their rights according to GDPR.

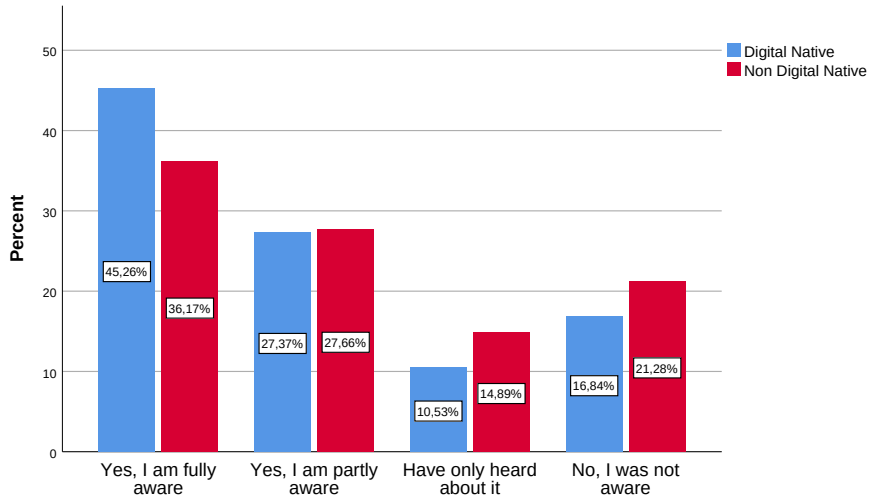


Figure 5.36: GDPR awareness of the control group respondents.

There was not found any significant difference between the digital natives and the non-digital natives concerning security awareness. However, the ANOVA one-way provided results indicating that there exist a trend between the groups. Digital natives seem to have a higher mean score overall throughout the questions regarding security awareness, as can be seen in figure 5.1.

Subsection:	Name:	Result:
Awareness about information gathering	Cookie awareness	ANOVA one-way: F(1, 134) = 6.489, p = .012
Awareness about information gathering	Data broker awareness	ANOVA one-way: F(1, 134) = 4.041, p = .046
Awareness about information gathering	Web beacon awareness	ANOVA one-way: F(1, 134) = 3.625, p = .059
Awareness about user's rights online	GDPR awareness	ANOVA one-way: F(1, 134) = 1.850, p= .176

Table 5.1: ANOVA one-way results for control group within Security awareness about information gathering

5.3.4 Risk perception and willingness

Between the digital natives and the non-digital natives, there was found significant differences in some of the scenarios:

Scenario 1, register a user on a social networking site: ANOVA one-way: F(1, 134) = 17,164, p <.001 showed that only digital natives was with 22,92% "very willing" to register a user. Further, digital natives also have a higher representative of responses with 55,21% in the willing (n=53) category. On the other hand,

non-digital natives are almost evenly spread between unwilling 38,3% (n=17) an willing 55,32% (n=20), with 6,38% (n=3) answers on very unwilling.

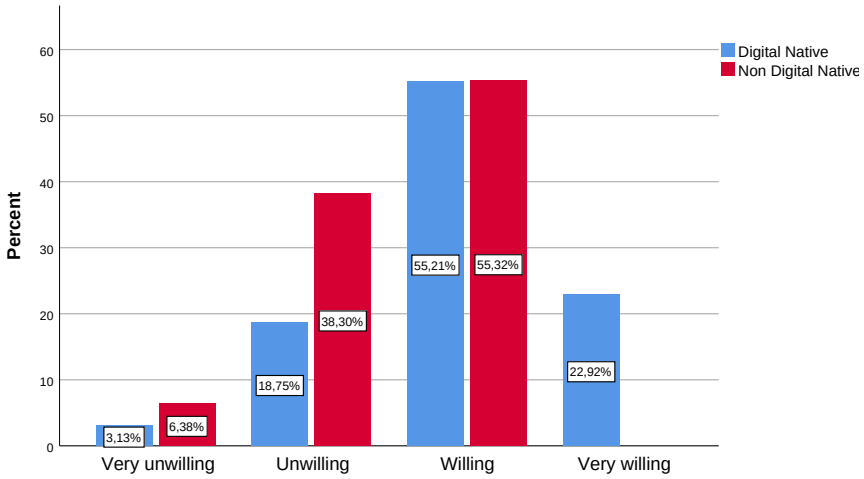


Figure 5.37: Comparison of age groups based on scenario 1 of all respondents.

Scenario 2, create user on a online newspaper: ANOVA one-way: $F(1, 134) = 5.970$, $p = .016$. The non-digital natives repsondents had 17,02% that were the "very unwilling" and the majority of answers were "unwilling" with 42,55%. Further, 31,91% were "willing", but only 8,51% were "very willing".

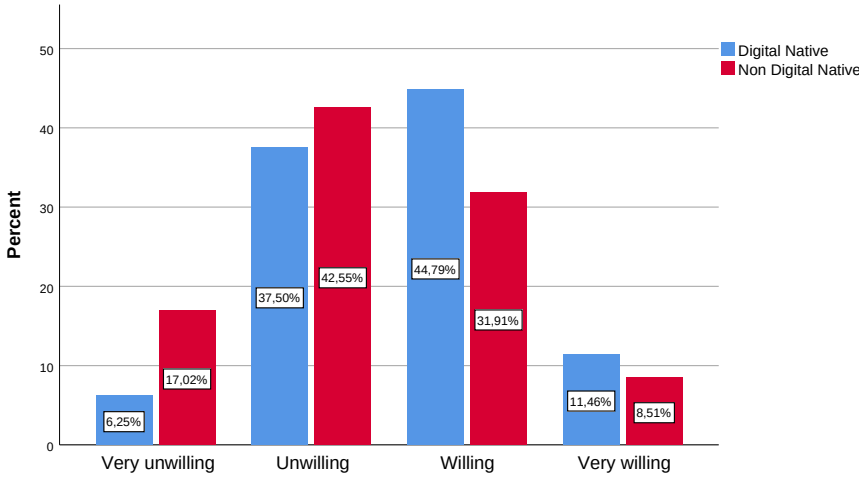


Figure 5.38: Comparison of age groups based on scenario 2 of all respondents.

Scenario 3, participate in a give away or survey on a blogg: ANOVA one-way: $F(1, 134) = 1.909$, $p = .169$. The vast majority of non digital natives were very

unwilling with 63,83%, and 29,79% were unwilling. On the other end, only 4,26% were willing and 2,13% very willing.

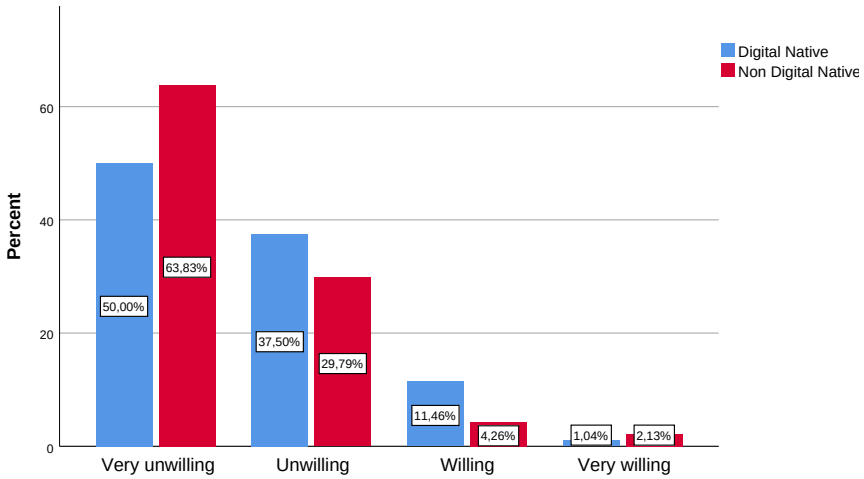


Figure 5.39: Comparison of age groups based on scenario 3 of all respondents.

Scenario4, shop for goods in a known online store. ANOVA one-way: $F(1, 134) = 23.109$, $p < .001$ shows that 10,64% the non digital are very willing 10, while 51,06% are willing. Further, 23,4% of the non digital natives are unwilling and 14,89% very unwilling. A comparison to the digital natives, shows that the non digital natives scores are less than the digital natives.

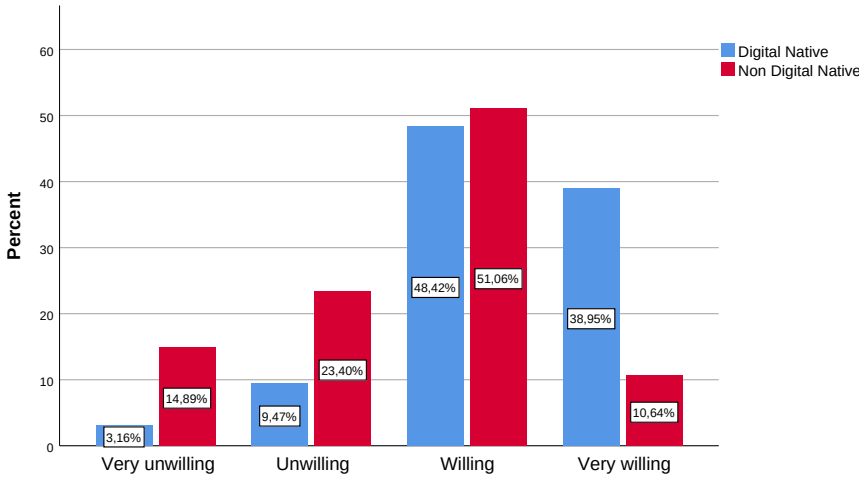


Figure 5.40: Comparison of age groups based on scenario 4 of all respondents.

Scenario5, use services that registrer user's geolocation. ANOVA one-way: $F(1, 134) = 25.179$, $p < .001$ shows that only a few with 6,38% of the non digital are very willing, while 27,66 % are willing, compared to digital natives where 25% were very willing, and the majority with 48,96% being willing. Further, the majority with 48,94% of the non digital natives are unwilling, and 17,02% very unwilling.

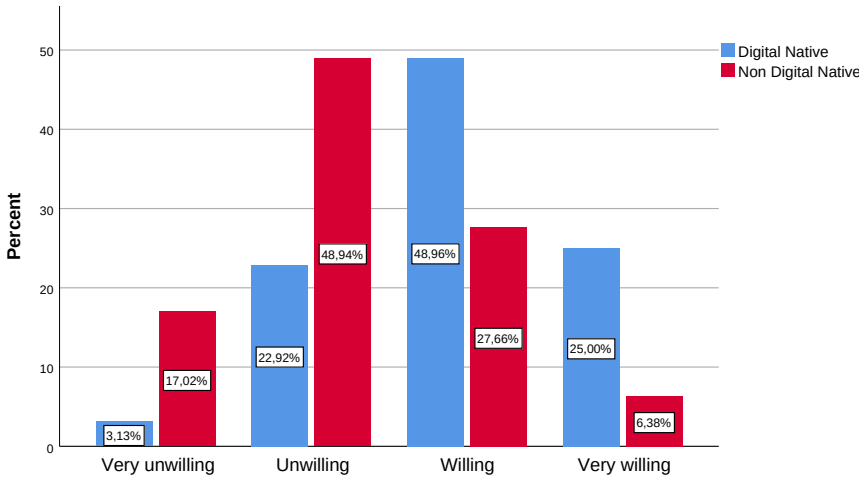


Figure 5.41: Comparison of age groups based on scenario 5 of all respondents.

Scenario 6, participate in a debate on a online forum: ANOVA one-way: $F(1, 134) = 4.983$, $p = .027$. None of the non digital natives are very willing, and only 2,13% are willing. The majority of the non digital native's responses are split between unwilling with 48,94% and very unwilling with 48,94%.

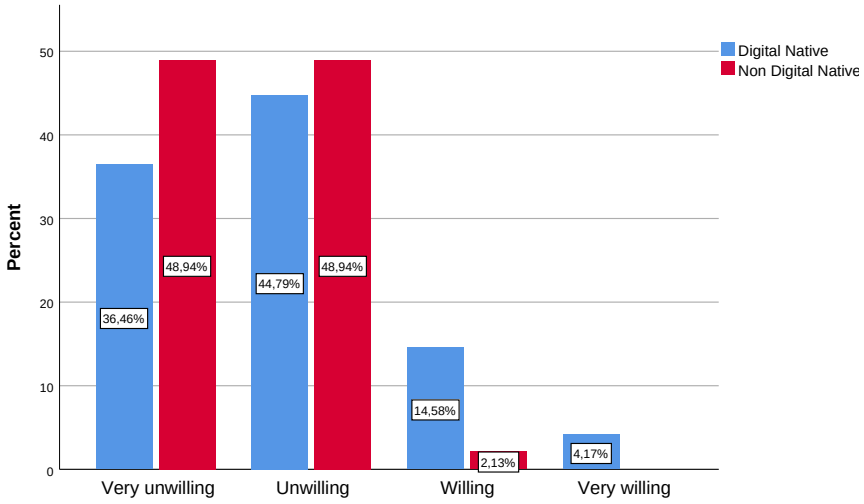


Figure 5.42: Comparison of age groups based on scenario 6 of all respondents.

There were found significant differences between the digital and non-digital natives concerning risk perception and acceptance. However, the significant difference did not occour for all of the scenarios. The ANOVA one-way results are summarized in figure 5.2 to provide an overview of the differences.

Subsection	Name	Result
Before accessing websites	Read policy before accessing websites	ANOVA one-way: $F(1, 134) = 5.239$, $p = .024$
Willingness to accept risk	Scenario 1	ANOVA one-way: $F(1, 134) = 17.164$, $p < .001$
Willingness to accept risk	Scenario 2	ANOVA one-way: $F(1, 134) = 5.970$, $p = .016$
Willingness to accept risk	Scenario 3	ANOVA one-way: $F(1, 134) = 1.909$, $p = .169$
Willingness to accept risk	Scenario 4	ANOVA one-way: $F(1, 134) = 23.109$, $p < .001$
Willingness to accept risk	Scenario 5	ANOVA one-way: $F(1, 134) = 25.179$, $p < .001$
Willingness to accept risk	Scenario 6	ANOVA one-way: $F(1, 134) = 4.983$, $p = .027$

Table 5.2: ANOVA one-way results for control group within risk perception and willingness

Chapter 6

Discussion

This chapter contains discussions and interpretations of the results from the previous chapter. Further, the primary purpose of the discussion chapter was to interpret how the results would contribute towards the conclusion. First, each research question was discussed, before presenting the strengths and weaknesses of the thesis.

Before discussing the research questions individually, it is beneficial to discuss the samples from both the content analysis and for both of the questionnaires. The samples are discussed in terms of rightful representation, size, and compared to other statistics. These factors will point towards strengths and weaknesses with the results and what to keep in mind when reading the discussion part, which will later be discussed more in-depth.

When examining specific groups such as digital natives, sample control must be emphasized to assure the sample consists of the intended group described in the methodology. In the background chapter 2.2.1, it was made clear that the definition of digital natives depends on two factors: age and raised in the digital era. Assuring the correct age was done by making the respondents select their age in the questionnaire and not including the answer submitted by respondents born before 1980. A study done by Eurostudent in 2018 [74], showed that 52% of the students in Norway were under 25 years old, but also that every fourth student was older than 30. Creating a small margin of error of representatives of older students in this study. The reason for zero respondents within the age group of 19 or younger, is explained by having students as the sample. In Norway, students start higher educations at universities at the age of 19 or older. Lastly, for the digital age it was fair to assume that students at NTNU have grown up with access to technology during the digital era.

Looking at the results from the questionnaire towards digital natives, the distribution of respondent's highest achieved education level seems to be reasonably representative compared to data collected from SSB. The data from SSB indicate

that 16% of the digital natives in Norway have a degree that takes more than 4 years to achieve, 36% have a degree until 4 year to achieve, 45% with videregående and 3% with fagskole [75] The comparison is illustrated in figure 6.1 and 6.2 below.

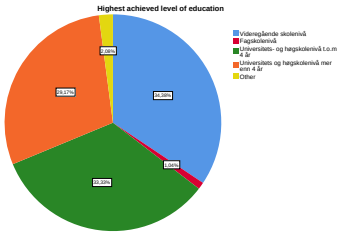


Figure 6.1: Questionnaire's highest achieved education for digital natives

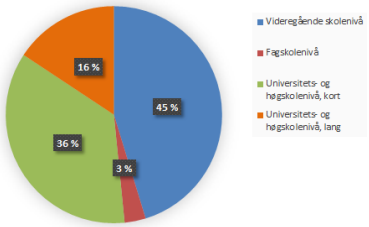


Figure 6.2: Norway's highest achieved level of education for digital natives

For gender, In Norway, SSB's data indicates that there are more women than men in higher educations, with 40,3% men and 59,7% women. In the original questionnaire towards digital natives, the men were higher represented with 58,3% and 41,67% women. This is not precisely like the statistics from SSB indicated, but close to a 50% distribution making the results a fair representative for both genders of the digital natives. In the control group, there was a larger representation of women respondents with 77,50% and 22,5% men. The over-representation of women is something to have in mind, but the results should still be able to raise arguments on behalf of non-digital natives. Lastly, there was a somewhat low amount of respondents in the both the questionnaires with N=96 from the digital natives and N=40 from the non digital natives. Keep in mind, that even if some parts of the study cannot be considered a statistically representative of the population, arguments and interpretations can still be valid to draw conclusions from and recommend further exploration in future work.

6.1 RQ1: How do information gathered differ depending on the business model of the website?

The primary outcome for research question 1, was to establish a foundation and provide results for use later in the thesis. As described in the methodology, the method to examine research question 1 was based on a paper by Pollach, I. [46] called "*What's wrong with online privacy policies?*". The method tried to answer how data is handled by reading the privacy policy provided to the users, using the "at-least-some" rule, resulting in an answer of "yes", even if the policy stated occasionally. This generalized the answers from policies and made a website collecting data one time and a website simultaneously collecting the data coded equally in the results. However, since research question 1 was intended to provide results for the upcoming questionnaire and not explicitly establish a deep insight into

the policies, the method was suitable for this purpose. In the paper by Pollach, I. the policies in their sample were coded twice, this increased the reliability of the data. The policy answers were only coded once in this thesis, because of the resources available, such as time and personnel. This might reduce the reliability, but by following the method, the results can still produce the intended content. For future work, the coding part of the data could be improved with more resources available, or other methods for producing the same results in a different matter. A suggestion considered was a method using a goal taxonomy for comparing privacy-policy statements into two categories, either privacy protection or privacy vulnerability as seen in the study by Earp, J, et. al. [44]. The method seen in Earp's paper was somewhat similar to the one used in this and could produce the wanted results. However, the choice to stick with pollach's method was ultimately decided on the taxonomy used in paper found more suited for this thesis.

Based on the empirical data from the content analysis, results from the sample in this thesis with 20 well-known websites indicate that all websites collect PII about their users, as well as aggregate information. Compared to the results from Pollach's original study, the number of questions that could not be answered decreased from 39,4% to 13,09%, resulting in a 26,31% increase of questions to be answered. This can be explained by the original study from 2007 was conducted before the implementation of GDPR in Europe. In 2018, Norway was bound to implement GDPR as a result of their membership in EØS. Providing more detailed privacy policies based on requirements from the law to fulfill. This also explains why every policy in the content analysis mentioned user's rights in regards to GDPR.

As a result of this information gathered about the different business models, the content analysis provided empirical data to construct specific questions to use in the questionnaires. Based on the taxonomy 5.1 and the results described above, and as seen in the table 6.1 below, the follow questions were constructed:

Area:	Question:
Security Awareness	Cookie awareness
Security Awareness	Data broker awareness
Security Awareness	Web beacon awareness
Security Awareness	GDPR Awareness
Risk Acceptance	Cookie acceptance
Risk Acceptance	Read privacy policies
Risk Perception	Scenarios 1-6

Table 6.1: Questions constructed based on the content analysis

The result from the content analysis provided the questionnaire with data to construct questions. The content analysis required time and research to figure out how to extract information out of policies. A more natural way to provide data to the questionnaire could be by looking at related work for what information other research had collected and phrase questions in a similar way. However, since the data privacy landscape is always changing with new laws implemented, both companies and users getting more aware of their responsibilities and rights concerning data privacy. The amount of related work available that fulfills the requirements for this thesis was limited. For example, older surveys do not provide the same valid results after the implementation of GDPR, and the websites used in the sample must be related to websites the digital natives in Norway use daily. Therefore, it was better to construct a research question and conduct a method in order to achieve and fulfill the requirements. The content analysis provided the results that were wanted but was more time consuming than first assumed. Results from the content analysis provided what the current privacy policies describe and made the content more relatable to the respondents in the questionnaire.

6.2 RQ2: To what degree are digital natives aware of the information gathering about their data when browsing the internet?

The questionnaire covered three terms related to data privacy and one question about the users' right online to measure awareness when browsing the internet. First, the answers from the original questionnaire towards the digital natives was discussed before compared to the control group of non-digital natives. These res-

ults, together with related work, will contribute towards a conclusion for the hypothesis.

6.2.1 Hypothesis 1: Higher education level achieved will result in an increased awareness for digital natives when browsing the internet.

Empirical data from the questionnaire, as seen in figure 5.5, shows that the higher the achieved level of education, the fewer people with no information security experience. This indicates that the more years digital natives spend studying towards a longer education, they get information security experience. However, when looking at the faculty distribution from the sample with digital natives in this questionnaire, there's a predominance of students from the faculty IE. The studies within this faculty are more likely to include electives containing information security because of the relevance of security in information technology and electrical engineering.

To further analyze if education level affects the degree of awareness for digital natives, the results from the questionnaire indicate that there is no significant effect on higher levels of education and awareness of terms related to information gathering. Even though, the mean value for the different education levels seems to be lower for respondents with higher education *(remember: 1 = fully aware, 2 = partly aware, 3 = heard about it, 4 = not aware)*. This seems to be a reoccurring trend for the terms "cookie" and "web beacon" asked in the questionnaire, as seen in figure 5.13. Indicating that education level might have a small, but no significant influence on the awareness of information gathering terms for digital natives. On the other hand, for the term "data broker", the mean value lowers from 2.70 for videregående education to 2.06 for the university until 4. years and back up to 2.11 for university longer than 4 years. This is making the results vary between the terms, ultimately resulting in no significant effect. However, the mean value decreases from videregående education to university education longer than 4 years for all the terms, indicating that higher education has a effect on awareness about data privacy terms.

The results were in line with previous studies such as the security awareness of digital natives written by Gkioulos, V. et al. [58]. Digital native's security awareness on specific areas was not significantly affected by their background. These results should be taken into account when considering the results from Gunleifsen, H. [56], where there was found that educated people in Norway consider themselves to know more. Based on the discussion, there was no significant evidence that higher education increased the awareness of digital natives. Some differences between the respondents from the education levels were found, raising an argument that the awareness about information gathering slightly increases along with education level. However, the results were not consistent for all of the questions,

but it can be interesting for future work to test on a larger scale with bigger sample size.

6.2.2 Hypothesis 2: Digital Natives are more aware of information gathering when browsing the internet than non-digital natives.

It was beneficial to examine how the digital natives' awareness score was in the questionnaire, then compare it to other sources to answer hypothesis 2. In addition to age, the number of hours spent online was significantly different between the digital natives and the non-digital natives. While the number of hours spent online did have some relation with the effect on the awareness of the digital natives, the results were not consistent and not will not be further discussed. For the awareness about the term "cookies", the results indicate that digital natives are very aware of this term. In total, 96,87% were fully or partly aware of cookies and how they are used on the internet to gather information. Less than 3% was not aware, reinforcing the argument for a high awareness regarding cookies for digital natives. Compared to the non-digital natives in the control group, they also have a majority of the respondents within the fully aware and partly aware category. However, what distinguishes the results, is that digital natives have 70,04% of the respondents who are fully aware, compared to the 55,32% from the non-digital natives. On the other hand, the non-digital natives have more respondents in the partly aware category, with 36,17% compared to 20,83% from the digital natives. Indicating that digital natives have a slightly higher awareness about cookies and their function.

Secondly, looking at the results for the term "data broker". A term that has been getting more attention after recent data privacy breaches such as the Facebook and Cambridge Analytica scandal in 2018 and the Netflix movie (the great hack). Even with the recent coverage, 22,92% of the digital natives were still not aware of the term. As mentioned in the background chapter 2.2.2, these companies have a significant impact on how the user's data is sold. Since these companies have a major influence on data privacy, digital natives must understand the concept of data brokers. The measurement of their knowledge was used as one of the reference points to measure the degree of awareness of information gathering. Of the digital natives, 61.46% were fully aware or partly aware, indicating that over half of the digital natives have awareness regarding data brokers. When comparing the results to the non-digital natives, there was a larger part of responses within the not aware category. 34,04% of non-digital natives are not aware of what a data broker is, compared to the 22,92% of digital natives. There was also a larger percentage of non-digital natives that only have heard of data brokers with 17,02% compared to the 15,63% from the digital natives. Furthermore, the digital natives score higher on both fully aware and partly aware than non-digital natives, indicating that digital natives have a higher awareness about the term data broker than non-digital natives.

Thirdly, for the measurement of awareness of terms. A more commonly unknown function implemented on websites known as "web beacons" was described to the respondents in the questionnaire to check their awareness. As described in the background chapter 2.2.2, these beacons are used to monitor the user's interactions on the particular website for information gathering. Awareness of these beacons is essential for users to understand what information websites gather about them. Looking at the results from the digital natives, the general awareness of web beacons was lower than the previous terms in the questionnaire. 20,83% were aware, 18,75% partly aware, and 18,75% have only heard about it, compared to the majority of 41,67% that were not aware. Having less % of digital natives aware or partly aware than the % of not aware indicates that the digital native's awareness was not the highest concerning web beacons. When compared with the non-digital natives, the non-digital natives had a higher percentage, with 53,19% were not aware of web beacons. That is 11,52% more answers in the same category, raising an argument that digital natives have a higher awareness of the web beacons. For the rest of the categories, non-digital natives scored lower in both fully aware and partly aware, and 3% than the digital natives only heard about it. The distribution, as seen in figure 5.35, produced a somewhat similar trend but raised an argument for an indication of higher awareness within digital natives.

To examine how the digital natives and the non-digital natives were aware of their rights when browsing the internet, one of the questions in the questionnaire was focused on the new law "GDPR" implemented in 2018 in Europe. The results indicated that the digital natives are aware of this data privacy law and what it contains, with 45,26% fully aware and 27,37% partly aware. Further, the results from the digital natives show that only 16,84% were unaware of their rights stated in GDPR. This can be explained by the heavy focus on data privacy and the media coverage of GDPR in Norway. Nevertheless, this does not change the fact that digital natives were updated on their rights online. Comparing the results to the non-digital natives showed that a general awareness also existed in the control group sample. Figure 5.36 illustrates that the distribution of awareness seems to be roughly the same, continuing the trend from the other questions regarding data privacy awareness. For GDPR awareness, the non-digital natives had a noticeable lower percentage of answers in the category fully aware, as well as a higher percentage in "I was not aware" category.

Digital natives tend to have an overall higher awareness regarding the three terms closely related to data privacy; this also accounts for the GDPR question about user's rights. The results from the questionnaire and the control group lead towards a slightly higher awareness for digital natives in regards to non-digital natives. These results build on existing evidence of the study by Gunleifsen, H. [56], where young people scored themselves relatively higher than the elderly.

6.3 RQ3: To what degree do digital natives accept risk and provide information?

Research question 3, involved measuring the willingness for digital natives to accept risks to access benefits websites can provide. For the hypotheses within research question 3, the discussion was based on the part of the questionnaire regarding risk perception. Firstly, towards digital natives, then compared with the control group. Before interpreting related work to the findings, ultimately resulting in a conclusion for each of the hypotheses.

6.3.1 hypothesis 1: Higher education level achieved decreases the willingness of digital natives to provide information.

To discuss hypothesis 1, the results from particularly 2 questions from the questionnaire were essential. The choices for users to opt-out and share less information happens before accessing the websites. The information about how to limit the information websites gather and what information the websites collect are found in the privacy policies. These policies are usually referred to on the cookie warning pop-up when entering the website. The 2 questions in the questionnaire related to "before accessing the website" were: cookie acceptance and if the respondents read the policy before accessing the website.

Looking at figure 5.15, the results show that the vast majority always accept all cookies when entering a website. On the other hand, 34,38% only keep strictly necessary cookies. These results indicate that the acceptance level was already high among the digital natives, where almost 2/3 of the respondents accepted all use of cookies. The ANOVA one-way test did not show any significant effect between higher education level and the acceptance of cookies. Indicating that a generally higher level of education and the willingness to accept cookies have no relationship.

Moreover, for the other question regarding before accessing the website, "do you read the privacy policy before accessing a website?". The results indicated that a majority of the digital natives never read the privacy policy before accessing the website. However, 38,54% sometimes read it. When conducting an ANOVA one-way, the p-value returns as .008, which is not significant but indicate some differences between the different education levels. These differences are further examined through the Tukey post hoc test. This test returned values showing educational level providing a difference in if the privacy policy gets real. However, a higher achieved educational level does not appear to significantly increase how many users read the policy before accessing the website.

From the discussion of hypothesis 1, a higher level of education did not have a significant effect on the willingness to provide information. The majority of digital

natives accepted cookies independent of their educational background. Furthermore, the results indicate that higher education made the digital natives more likely to sometimes read the policy before accessing a website. However, these results were not significant but raised an argument for further work.

6.3.2 hypothesis 2: Digital Natives are less likely to accept risks than non digital natives

The willingness to accepting risks among digital natives and compare it to non-digital natives was examined through hypothesis 2. In both questionnaires, there were 6 questions phrased as scenarios for the respondents to answer their degree of willingness for each specific scenario. In order to discuss the results for hypothesis 1, a closer look at the results from the scenarios was beneficial, then compare with the results from the non-digital natives and related work.

Scenario 1: the results showed that the digital natives had a higher willingness to register a user at online networking sites. By comparing the results of the digital natives with the non-digital natives, the results point towards digital natives being more willing to accept the risk for direct marketing to achieve the benefit of interacting on a social networking site

Scenario 2: The results from the questionnaires indicated that the willingness to register a user on online newspapers is close to evenly distributed between the digital natives and the non-digital natives. By looking at figure 5.38, the results showed that the digital natives have a total of 15,83% more in the willing and very willing categories. However, this does not result in a significant difference between the group, where the ANOVA one-way gave a p-value of .016, indicating that the digital natives are a bit more willing than the non-digital natives.

Scenario 3: examining the distribution from figure 5.39, showed that there was no significant difference in participating in a giveaway or survey on blogs between the digital natives and the non-digital natives. However, digital natives had a slightly higher representation in the two categories willing and very willing, as seen in previous scenarios. Indicating that there is no significant difference, but the trend is pointing towards higher willingness for digital natives compared to non-digital natives.

Scenario 4: the results indicated a difference between the groups. The ANOVA one-way showed a p-value $<.001$, which means that there were significant differences in the responses from digital natives and non-digital natives for shopping goods online from web stores. Looking at the figure 5.40, as for the previous scenarios, the digital natives scored higher in the awareness categories, than the older generations. The vast majority of digital natives with 87,37% had answered willing or very willing, compared to 61,70% from the non-digital natives. Raising an

argument for the digital natives being more willing to prioritize access over the risk.

Scenario 5: the use of services with the users' geolocation. The results showed a significant difference between the groups, with ANOVA one-way giving a p-value of <.001. By further examining the figure 5.41, the distribution indicated that the digital natives had a majority of respondents in the willing and very willing categories. While the non-digital natives were represented by a vast majority of answers in the unwilling category. Furthermore, the Non-digital natives had 13,89% higher representation in the very unwilling category. Results from both questionnaires indicated that digital natives were significantly more willing to accept risk for access in scenario 5.

Scenario 6: the willingness to participate in a debate on an online forum. Looking at the distribution from figure 5.42, non-digital natives had none respondents that were "very willing" and only 2,13% "willing". Compared to digital natives that have 14,58% "willing" and 4,17% "very willing". ANOVA one-way gave a p-value of .027, indicating no significant difference. However, non-digital natives had more respondents in the "very unwilling" and "unwilling" categories than the digital natives, raising an argument that the digital natives were more willing to accept risk for access in scenario 6.

The trend for digital natives seems to be more willing to accept risk than non-digital natives. Through the scenarios, the digital natives had slightly or significantly more answers in the "willing" and "very willing" categories compared to the non-digital natives. Out of 6 scenarios, 3 of them resulted in a p-value <.001, providing a clear significance between the groups. Based on the empirical data from the questionnaire towards both groups, the results lead to a rejection of hypothesis 1. Instead, digital natives seemed to be more willing to accept risk than non-digital natives. The results do not fit with the report "The Norwegian Cyber Security Culture" published by NorSiS[63] in 2019, where most of the population in Norway were worried about their data being collected by companies online. Indicating that digital natives might be more willing to accept risks related to specific scenarios than when asked about general collection about their data.

6.4 Strength and limitations

To reflect on the work done in the thesis, it is beneficial to look at the thesis's strengths and weaknesses.

6.4.1 Strengths

With the use of mixed methods, collection of data happened through both a qualitative method and then a quantitative method. Mixed methods utilize the strengths

of both the methods, by doing a more in-depth analysis of the privacy policies and then using quantitative methods to get a larger picture. The results of this method gave a current, up-to-date insight into the content of these policies, which was then used in the questionnaires towards the target group, resulting in a deeper understanding of the security awareness and risk perception of digital natives Norway.

Further, the method for data collection produced the data as indented. The method used to collect data measured the variables and provided suitable and robust data for a thorough analysis of the results. Precautions were taken to ensure sample control, with a focus on reliability and validity. Ultimately, the method effectuated the data collection, which was the most crucial part of the study.

Lastly, the thesis was written during a time where awareness regarding data privacy is heavily sought after. The results produced in this thesis provide a scientific contribution to the field of data privacy. Either by using the thesis as a reference or as a basis to produce future work with new methods.

6.4.2 Limitations

As with the majority of studies, the design of the current study is subject to limitations. Through the data collection from the quantitative method, the sample of digital natives chosen in this thesis consisted of students at NTNU. Even though students represent a variety of backgrounds and qualifications, the sample used in this thesis does not cover the digital natives without any education. Moreover, the sample size is not found statistically representative for either the digital natives nor the non-digital natives. However, as previously mentioned, the results still gave an extensive and reliable overview of the security awareness and risk perception of the digital natives.

The time perspective is crucial for all scientific research. This thesis, a cross-sectional study, the data were collected one time. A potential limitation regarding the one-time collection of data is that there was only one sample to analyze. The validity of the results would increase by conducting the same study again on a later stage, either on the same sample or a new sample of digital natives.

As mentioned throughout the thesis, data privacy is still a relatively small field with limited publications, especially towards specific target groups. This results in few other related work publications for comparison of this study. The lack of previous research studies on the field limits the possibility of measuring the results toward similar data from multiple sources.

Chapter 7

Conclusion

7.1 Conclusion

The purpose of this thesis was to examine the digital native's security awareness and risk acceptance regarding data privacy. This involved a two-step process to produce the intended results. First, a content analysis was conducted on a sample of websites. The content analysis conducted on these websites provided results and content for the second process and the primary method of the thesis: a questionnaire towards the digital natives in Norway. The questionnaire was sent to digital natives to measure their awareness and risk perception through questions about data privacy. Lastly, the same questionnaire was sent out to non-digital natives as a control group to provide answers for comparison and answer the research questions.

Regarding security awareness, digital natives have a good awareness of the basic terms related to data privacy. However, the degree of awareness decrease as the terms get more technical. Further, regarding security awareness, the results indicate that digital natives also have a high awareness concerning their rights in relation to GDPR. There were some significant differences between the groups on certain types of questions regarding security awareness, but nothing consistent.

For risk perception, digital natives are likely to accept risks to achieve access to websites. The results from the scenarios indicated a clear trend, and found a significant difference between the groups. This significance indicated that digital natives have a high degree of risk acceptance related to risks that can threaten data privacy. Further, regarding the risk acceptance, digital natives show a high degree of willingness to accept cookies and not read the privacy policies. However, there was not found any significant difference found between the groups before accessing the websites.

These conclusions were established based on the collected data, related work, and the discussion regarding the research questions on security awareness and

the risk acceptance of the digital natives. Overall, there is a large potential to increase both the degree of awareness and the degree of risk perception among digital natives regarding data privacy. The results of this study open multiple paths for future work as further described in the next chapter.

Chapter 8

Future Work

This chapter contains a discussion regarding recommendations for future work. The chapter is broken into sub-chapters with a different perspective for recommending future studies based on this thesis.

8.1 Recommendations

8.1.1 Should further research be conducted towards the same research questions?

The research questions are worth further research in future studies. A focus on an even broader and larger sample of digital natives can create a better picture of the situation regarding digital natives and data privacy awareness and risk perception. As data privacy is a field in constant development and subject to change. It would be beneficial to repeat the study to see if there is any development in the degree of security awareness and risk perception regarding data privacy among digital natives. as well as other what other laws that influence the every day of digital natives.

8.1.2 Should the research be repeated with other methods?

For future work, it can be valuable to look at different ways to measure awareness, including new terms of both more and less technical. A qualitative method towards the digital natives' degree of security awareness and risk perception could be conducted to get a more in-depth understanding of the degree. A questionnaire has its pros and cons; for example, the questions can be interpreted in different ways without the possibility to ask questions towards the interviewer. By conducting interviews or using observation, future studies might get more detailed answers which can be compared to this study.

8.1.3 Is it necessary to go more in-depth on certain areas?

Moreover, since this thesis looked into the degree of security awareness and risk perception, a study to explore further what the consequences have been for breaches of privacy and what measures can decrease and limit the outcome of the consequence. For example, conducting a qualitative research study towards those that have been a victim of a privacy breach. This can provide recommendations for how digital natives can improve their security and raise their awareness related to data privacy.

8.1.4 Have the research resulted in new topics that should be explored?

Through the research conducted towards digital natives, there have emerged new topics that can be interesting to explore. For example, the results indicate that the number of hours spent online browsing the internet seem to affect parts of the awareness and risk perception. In future study, it can be interesting to further look into how more time interacting on the internet increase security awareness or risk perception. Furthermore, digital natives seem to have a good understanding of their rights in relation to GDPR. On the other hand, it could be interesting to examine the companies understanding of their rights and responsibilities in relation to data privacy. Lastly, more general research towards the whole population's data privacy could be valuable to examine the landscape of data privacy further.

Bibliography

[1] F.-o. administrasjonsdepartementet, *Hva er personvern?*, Library Catalog: www.regjeringen.no, Oct. 2019. [Online]. Available: https://www.regjeringen.no/no/tema/statlig-forvaltning/personvern/hva-er-personvern/id448290/.

[2] Library Catalog: www.merriam-webster.com. [Online]. Available: https://www.merriam-webster.com/dictionary/privacy.

[3] Library Catalog: www.businessdictionary.com. [Online]. Available: http://www.businessdictionary.com/definition/privacy.html.

[4] R. Gavison, 'Privacy and the limits of law', *The Yale Law Journal*, vol. 89, no. 3, pp. 421–471, 1980.

[5] A. F. Westin, *Privacy and freedom*. Atheneum, 1967, ISBN: 978-1-935439-97-4.

[6] E. Shils, 'Privacy: Its constitution and vicissitudes', *Law and Contemporary Problems*, vol. 31, no. 2, pp. 281–306, 1966, ISSN: 00239186. [Online]. Available: http://www.jstor.org/stable/1190672.

[7] S. D. J, 'A taxonomy of privacy', *U. Pa. L. Rev.*, vol. 154, p. 477, 2005.

[8] J. R. Reidenberg, 'Resolving conflicting international data privacy rules in cyberspace', *Stanford Law Review*, pp. 1315–1371, 2000.

[9] *Fair information practice principles*, 2007. [Online]. Available: https://web.archive.org/web/20090331134113/http://www.ftc.gov/reports/privacy3/fairinfo.shtm.

[10] P. M. Schwartz and D. J. Solove, 'The pii problem: Privacy and a new concept of personally identifiable information', *NYUL rev.*, vol. 86, p. 1814, 2011.

[11] *Protection of individuals with regard to the processing of personal data and on the free movement of such data*, 1995. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31995L0046.

[12] *Surveillance giants*, 2019. [Online]. Available: https://www.amnesty.org/en/documents/pol30/1404/2019/en/.

[13] A. Etzioni, 'Are new technologies the enemy of privacy?', *Knowledge, Technology & Policy*, vol. 20, pp. 115–119, 2007. DOI: 10.1007/s12130-007-9012-x.

[14] D. Solove, 'The end of privacy?', *Scientific American*, vol. 299, pp. 100–4, 106, 2008. DOI: 10.1038/scientificamerican0908-100.

[15] P. W. Watson, *This is the end of privacy as we know it*. [Online]. Available: https://www.forbes.com/sites/patrickwwatson/2018/04/26/this-is-the-end-of-privacy-as-we-know-it/.

[16] [Online]. Available: https://www.philosophytalk.org/blog/end-privacy.

[17] S. Chakravarty, *The end of privacy?*, Apr. 2019. [Online]. Available: https://www.geospatialworld.net/article/the-end-of-privacy/.

[18] A. Preston, 'The death of privacy', *The Observer*, Aug. 2014, ISSN: 0029-7712. [Online]. Available: https://www.theguardian.com/world/2014/aug/03/internet-death-privacy-google-facebook-alex-preston.

[19] C. J. Sykes, *The end of privacy*, 1st ed. St. Martin's Press, 1999, ISBN: 978-0-312-20350-4.

[20] *Harmful*. [Online]. Available: https://www.merriam-webster.com/dictionary/harmful.

[21] P. W. van den Hoven Jeroen Blaauw Martijn and W. Martijn, *Privacy and information technology*, E. N. Zalta, Ed., 2020. [Online]. Available: https://plato.stanford.edu/archives/sum2020/entries/it-privacy/.

[22] S. A. Vinterbo, *Views on privacy*, Sep. 2019.

[23] J. G. Palfrey and U. Gasser, *Born digital: Understanding the first generation of digital natives*. ReadHowYouWant.com, 2011.

[24] I. C. of Commerce, *Icc uk cookie guide*, 2019. [Online]. Available: https://ico.org.uk/for-organisations/guide-to-pecr/guidance-on-the-use-of-cookies-and-similar-technologies/.

[25] L. C. Management, *The world's most valuable resource is no longer oil, but data*, 2019. [Online]. Available: https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.

[26] A. Kuempel, 'The invisible middlemen: A critique and call for reform of the data broker industry', *Nw. J. Int'l L. & Bus.*, vol. 36, p. 207, 2016.

[27] F. Commission, 'Data brokers: A call for transparency and accountability', in. 2014, pp. 1–101.

[28] M. P. Bailey and B. M. Error, *Web usage overlays for third-party web plug-in content*, US Patent 7,584,435, Sep. 2009.

[29] A. R. A. Bouguettaya and M. Y. Eltoweissy, 'Privacy on the web: Facts, challenges, and solutions', *IEEE Security Privacy*, vol. 1, no. 6, pp. 40–49, 2003.

[30] 2018. [Online]. Available: https://ec.europa.eu/info/law/law-topic/data-protection/eu-data-protection-rules_en.

[31] *2018 reform of eu data protection rules*, European Commission, 25th May 2018. [Online]. Available: https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-changes_en.pdf.

[32] *Individual rights*, Information Commissioner's Office, 25th May 2018. [Online]. Available: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/.

[33] T. Busch and Fagbokforlaget, *Akademisk skriving for bachelor- og masterstudenter*. Fagbokforlaget, 2013, ISBN: 9788245014426. [Online]. Available: https://books.google.no/books?id=HuBFngEACAAJ.

[34] P. Leedy and J. Ormrod, *Practical Research: Planning and Design, Global Edition*. Pearson Education Limited, 2015, ISBN: 9781292095882. [Online]. Available: https://books.google.no/books?id=2v0wCwAAQBAJ.

[35] A. Tjora, *Kvalitative forskningsmetoder i praksis. 3. utgave.* 2017, ISBN: 978-82-05-50096-9.

[36] E. Bertsen K., *Vitenskapelig forankring for bacheloroppgave ved anvendt informasjonsvitenskap v18. institutt for datateknologi og informatikk*, NTNU, Jan. 2018.

[37] Hoofnagle, C. Jay and Nathan, *Web privacy census*, Jul. 2014. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2460547.

[38] A. Cahn, S. Alfeld, P. Barford and S. Muthukrishnan, 'An empirical study of web cookies', in *Proceedings of the 25th International Conference on World Wide Web*, ser. WWW '16, Montréal, Québec, Canada: International World Wide Web Conferences Steering Committee, 2016, pp. 891–901, ISBN: 9781450341431. DOI: 10.1145/2872427.2882991. [Online]. Available: https://doi.org/10.1145/2872427.2882991.

[39] *About cookiepedia*. [Online]. Available: https://cookiepedia.co.uk/about-cookiepedia.

[40] Soltani, Ashkan, Canty, Shannon, Quentin, Thomas, Lauren, Hoofnagle and C. Jay, *Flash cookies and privacy*, Aug. 2009. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1446862.

[41] M. D, Wambach, D. James, Soltani, Ashkan, Nathan, Hoofnagle and C. Jay, *Flash cookies and privacy ii: Now with html5 and etag respawning*, Jul. 2011. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1898390.

[42] A. Dabrowski, G. Merzdovnik, J. Ullrich, G. Sendera and E. Weippl, 'Measuring cookies and web privacy in a post-gdpr world', in *Passive and Active Measurement*, D. Choffnes and M. Barcellos, Eds., Cham: Springer International Publishing, 2019, pp. 258–270.

[43] P. T. AG, *Cookies, the gdpr, and the eprivacy directive*, May 2019. [Online]. Available: https://gdpr.eu/cookies/.

[44] J. Earp, A. Antón, L. Aiman-Smith and W. Stufflebeam, 'Examining internet privacy policies within the context of user privacy values', *Engineering Management, IEEE Transactions on*, vol. 52, pp. 227–237, Jun. 2005. DOI: 10.1109/TEM.2005.844927.

[45] A. I. Anton, J. B. Earp and A. Reese, 'Analyzing website privacy requirements using a privacy goal taxonomy', in *Proceedings IEEE Joint International Conference on Requirements Engineering*, Sep. 2002, pp. 23–31. DOI: 10.1109/ICRE.2002.1048502.

[46] I. Pollach, 'What's wrong with online privacy policies?', *Communications of the Association for Computing Machinery*, vol. 50, no. 9, pp. 103–108, 2007, ISSN: 0001-0782.

[47] A. D. Miyazaki, 'Online privacy and the disclosure of cookie use: Effects on consumer trust and anticipated patronage', *Journal of Public Policy & Marketing*, vol. 27, no. 1, pp. 19–33, 2008. DOI: 10.1509/jppm.27.1.19.

[48] A. M. McDonald, R. W. Reeder, P. G. Kelley and L. F. Cranor, 'A comparative study of online privacy policies and formats', in *Privacy Enhancing Technologies*, I. Goldberg and M. J. Atallah, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 37–55, ISBN: 978-3-642-03168-7.

[49] P. Kelley, J. Bresee, L. Cranor and R. Reeder, 'A "nutrition label" for privacy', Jan. 2009. DOI: 10.1145/1572532.1572538.

[50] M. W. Vail, J. B. Earp and A. I. AntÓn, 'An empirical study of consumer perceptions and comprehension of web site privacy policies', *IEEE Transactions on Engineering Management*, vol. 55, no. 3, pp. 442–454, Aug. 2008, ISSN: 1558-0040. DOI: 10.1109/TEM.2008.922634.

[51] M. S. Ackerman, L. F. Cranor and J. Reagle, 'Privacy in e-commerce: Examining user scenarios and privacy preferences', in *Proceedings of the 1st ACM conference on Electronic commerce*, 1999, pp. 1–8.

[52] A. I. Antón, J. B. Earp and J. D. Young, 'How internet users' privacy concerns have evolved since 2002', *IEEE Security & Privacy*, vol. 8, no. 1, pp. 21–27, 2010.

[53] B. Berendt, O. Günther and S. Spiekermann, 'Privacy in e-commerce: Stated preferences vs. actual behavior', *Communications of the ACM*, vol. 48, no. 4, pp. 101–106, 2005.

[54] N. Hoebel and D. Zumstein, 'Trust and privacy in the web: Do internet users feel monitored on websites?', *ACM Web Science 2011*, Jun. 2011.

[55] N. Gerber, V. Zimmermann and M. Volkamer, 'Why johnny fails to protect his privacy', in *2019 IEEE European Symposium on Security and Privacy Workshops (EuroS PW)*, Jun. 2019, pp. 109–118. DOI: 10.1109/EuroSPW.2019.00019.

[56] H. Gunleifsen, V. Gkioulos, G. B. Wangen, A. Shalaginov, M. Kianpour and M. Abomhara, 'Cybersecurity awareness and culture in rural norway', in *null*, Jul. 2019.

[57] D. Ariu, F. Bosco, V. Ferraris, P. Perri, G. Spolti, P. Stirparo, G. Vaciago and S. Zanero, *Security of the Digital Natives*, ID 2442037. May 2014. DOI: 10.2139/ssrn.2442037. [Online]. Available: https://papers.ssrn.com/abstract=2442037.

[58] V. Gkioulos, G. B. Wangen, S. Katsikas, G. Kavallieratos and P. Kotzanikolaou, 'Security awareness of the digital natives', *Information (Switzerland)*, vol. 8, Apr. 2017. DOI: 10.3390/info8020042.

[59] S. M. Forsythe and B. Shi, 'Consumer patronage and risk perceptions in internet shopping', *Journal of Business Research*, vol. 56, no. 11, pp. 867–875, 2003, Strategy in e-marketing, ISSN: 0148-2963. DOI: https://doi.org/10.1016/S0148-2963(01)00273-9. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0148296301002739.

[60] NorSIS, 'The Norwegian Cyber Security Culture', Norsk senter for informasjonssikring, Tech. Rep., 2016.

[61] NorSIS, 'Nordmenn og digital sikkerhetskultur', Norsk senter for informasjonssikring, Tech. Rep., 2017.

[62] NorSIS, 'Nordmenn og digital sikkerhetskultur', Norsk senter for informasjonssikring, Tech. Rep., 2018.

[63] NorSIS, 'Nordmenn og digital sikkerhetskultur', Norsk senter for informasjonssikring, Tech. Rep., 2019.

[64] L. M. Coventry, D. Jeske, J. M. Blythe, J. Turland and P. Briggs, 'Personality and social framing in privacy decision-making: A study on cookie acceptance', *Frontiers in Psychology*, vol. 7, p. 1341, 2016, ISSN: 1664-1078. DOI: 10.3389/fpsyg.2016.01341. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2016.01341.

[65] A. Adams, 'Users' perception of privacy in multimedia communication', 1999. DOI: 10.1145/632716.632752.

[66] J. Bhatia, T. D. Breaux, J. R. Reidenberg and T. B. Norton, 'A theory of vagueness and privacy risk perception', in *2016 IEEE 24th International Requirements Engineering Conference (RE)*, 2016, pp. 26–35.

[67] In. SAGE Publications, Inc., Jun. 2020. DOI: 10.4135/9781483398075. [Online]. Available: https://methods.sagepub.com/book/factorial-survey-experiments.

[68] H. Hibshi, T. D. Breaux and S. B. Broomell, 'Assessment of risk perception in security requirements composition', in *2015 IEEE 23rd International Requirements Engineering Conference (RE)*, 2015, pp. 146–155.

[69] P. Mayring, 'Qualitative content analysis', *A companion to qualitative research*, vol. 1, pp. 159–176, 2004.

[70] *Out of control - how consumers are exploited by the online advertising industry*, Forbrukerrådet, Jan. 2020.

[71] Apr. 2015. [Online]. Available: https://no.ehandel.com/artikler/norges-10-storste-nettbutikker/380694.

[72] *Lov om behandling av personopplysninger (personopplysningsloven) lov-2018-12-20-116*, 2018. [Online]. Available: https://lovdata.no/dokument/NL/lov/2018-06-15-38.

[73] G. Norman, 'Likert scales, levels of measurement and the "laws" of statistics', *Advances in Health Sciences Education*, vol. 15, pp. 625–632, 2010.

[74] G. C. Hauschildt Kristina Vögtle Eva Maria, *Social and economic conditions of student life in europe*, 2018. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31995L0046.

[75] SSB, 2019. [Online]. Available: https://www.ssb.no/utniv.

Appendix A

Additional Material

A.1 Questionnaire

A.1.1 Background information

A variety of browsers seem to be in use by the respondents. This question was phrased as a multiple choice question, encouraging the users to chose several browsers if they use more than one. Since most people have more than one device connected to the internet, the browser used can vary between devices. Google Chrome stand out as the most preferred browser with 83,3% (n=80) of the respondents using this browser. Next, Safari is the second most used browser with 31,2% (n=30), slighty more used than Firefox with 24% (n=23). Microsoft edge 6,2% (n=6) and Opera 2,1% (n=2) are less in use by the digital natives in this questionnaire. Lastly, zero of the respondents uses Internet explorer.

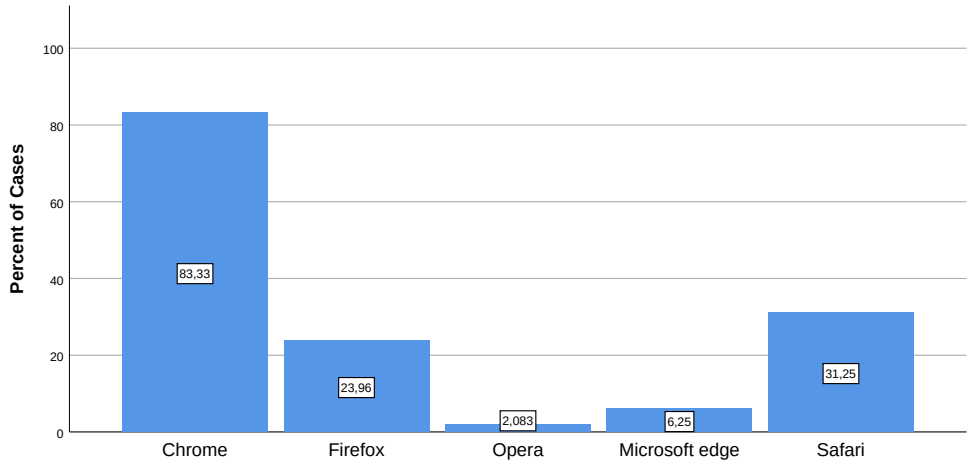


Figure A.1: Browsers used by the questionnaire respondents.

A.1.2 Measures taken

As the last question on the questionnaire, in order to measure how many of the respondents that have taken measures to protect their privacy online. The respondents were asked to check the answers if they had done the measure described, the alternatives can be seen in figure A.2. The results (N=92) show that the vast majority had done one or more measure to protect their privacy online, with only 4 people not selecting an alternative. The distribution shows that 60,87% had rejected the use of cookies, 69,57% had chosen not to use a website because of uncertainty concerning data collection. Further, the most selected answer with 72,83% was provided fake or fictive information under registration on a website, while 61,96% had requested one or more website to not share personal data with

third parties. Lastly, the least selected answer with 36,96% requested one or more website to delete all personal data stored about them.

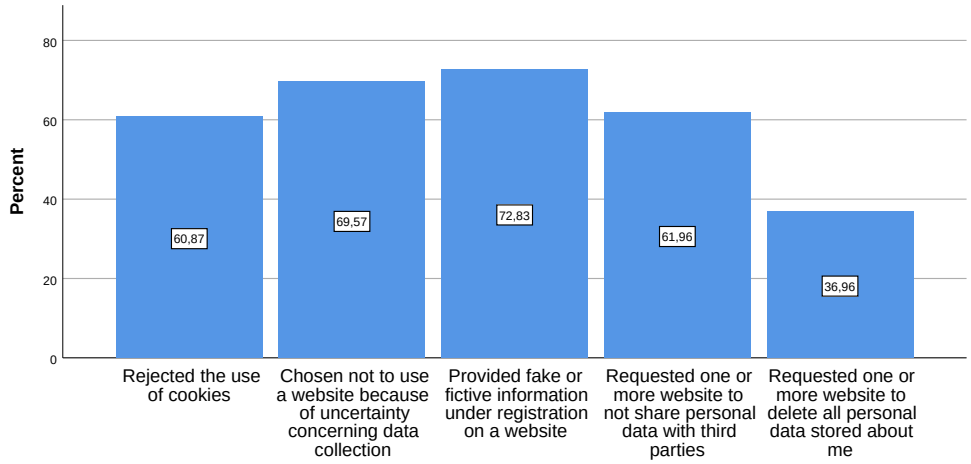


Figure A.2: Distribution of measures taken of the questionnaire respondents.

A.2 Control Group

A.2.1 Background information

For browser usage among the non digital natives, Chrome is the most used browser with 64,86% (n=24) of the cases, while Safari comes second with 56,8%. Further, Internet explorer is still used by the non digital natives with 40,54%(n=15), before Firefox with 13,51%(n=5) and Opera with only 2,7% (n=1).

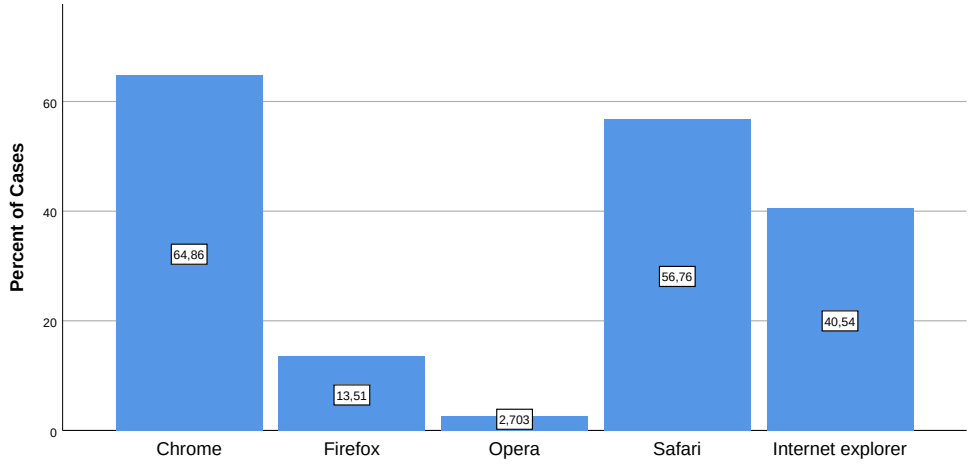
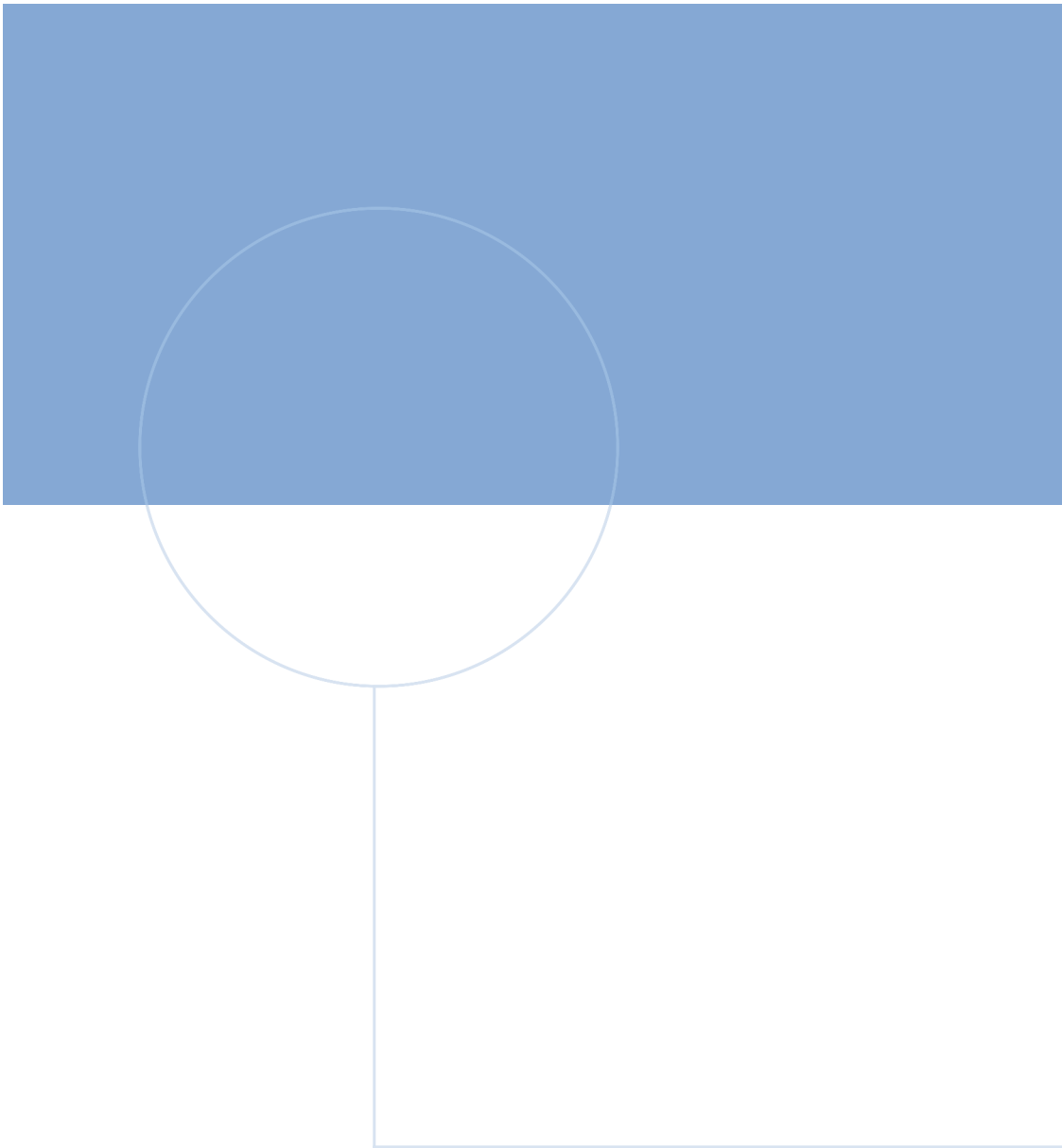


Figure A.3: Highest achieved level of education and no information security experience of the control group respondents.



NTNU

Norwegian University of
Science and Technology