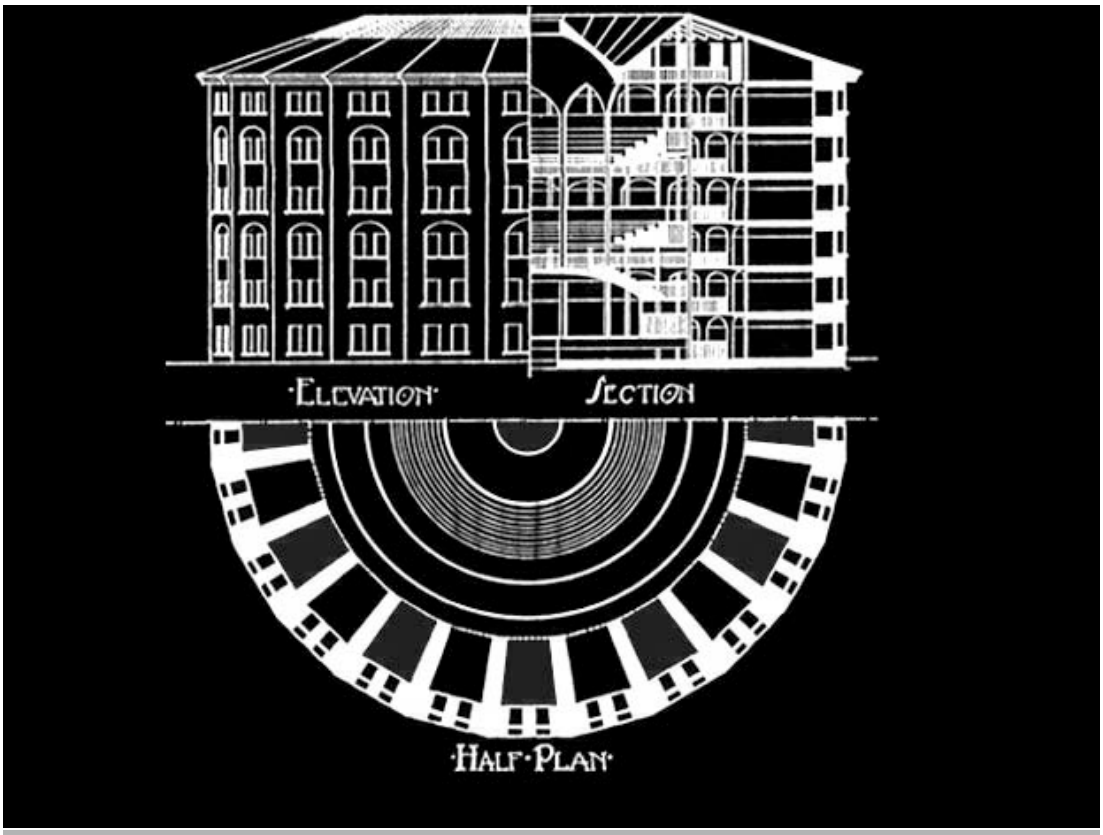
The Online Panopticon

Privacy Risks for Users of Social Network Sites



Identification and prioritization of privacy risks for users of Social Network Sites and considerations for policy makers to minimize these risks

David Riphagen

1102303

Systems Engineering, Policy Analysis and Management

Delft University of Technology

Version 1.0
Wednesday, October 8, 2008

Table of Contents

List of tables and illustrations **7**

Preface and acknowledgements **9**

Summary **11**

1. Social Network Sites: increasing in numbers and problems? **13**

1.1. Problem Analysis 13

1.2. Previous research 14

1.3. What this study achieves 15

1.4. Demarcation 16

1.5. How this study achieves this 17

2. What are Social Network Sites and why can they harm us? **21**

2.1. What is a Social Network Site? 21

2.2. What is identity relevant information and why is it important? 23

2.3. Reasons to protect users' information on Social Network Sites 25

2.4. How privacy risks materialize: risk analysis 30

3. Framework: structuring privacy harms in Social Network Sites **31**

3.1. Description of key factors of Social Network Sites 33

3.2. Classification of threats 34

3.3. Privacy incidents 35

3.4. Harmful outcomes or damages 35

3.5. Example of the use of the framework 36

4. Description of key factors of Social Network Sites **39**

4.1. Societal environment 39

4.2. Features of the medium relevant to SNS: Topology of SNS 50

4.3. Value chain and customers 53

4.4. Specific features of the product 56

4.5. Financial flow and flow of services 58

4.6. Conclusions 60

5. Classification of threats **61**

5.1. Information collection....61
5.2. Information processing....64
5.3. Information dissemination....68
5.4. Merging the privacy taxonomy with moral reasons to protect privacy....71
6. Privacy threats: acquiring input from privacy and Internet experts....73
6.1. Goals of the survey....73
6.2. Methodology....74
6.3. Process of surveying....76
6.4. Analysis of data....77
6.5. Priority setting and rater agreement measurement....77
6.6. Conclusions....82
7. Privacy threats, incidents and harms for users of SNS....83
7.1. Description of respondents....83
7.2. Privacy threats for users of SNS....84
7.3. Probability of occurrence and negative impact....86
7.4. Tort laws to minimize the damage from privacy threats....90
7.5. Facebook case study harms....94
7.6. Risks identified by the IWGDPT....94
7.7. Risks identified by ENISA....95
7.8. Conclusions....97
8. Examples of threats, incidents and damages....99
8.1. Total information awareness....99
8.2. Dissemination to wrongdoers....103
8.3. No control over information....108
8.4. Others....113
8.5. Conclusions....116
9. Reflections....117
9.1. Discussion on methodology....117
9.2. Discussing the substance....118
9.3. Discussion of the policy process....118

10. Conclusions and recommendations.....123

10.1. Conclusions.....123

10.2. Recommendations for further research and research agenda.....125

Glossary.....129

References.....135

Appendices.....149

List of tables and illustrations

Illustration 1: A rational model of policy formulation, adapted from Bots and Hulshof (1995)..... 18

Illustration 2: Research approach.....21

Illustration 3: Risk analysis framework and its relations with this study, as adapted from Van den Berg (2008).....32

Illustration 4: Beacon example analyzed with framework 40

Illustration 5: The MCM Framework used to describe Social Network Sites as adapted from Hoegg et al. (2006), with the different theories used to analyze the specific areas of the framework.....41

Illustration 6: The actor network of Social Network Sites, depicting the various actors that are part of the value chain.....56

Table 1: Options with pay-offs for Social Network Sites and users.....57

Illustration 7: Adapted version of the Privacy taxonomy of Solove (2006).....63

Table 2: Correlation between Solove's (2006) Privacy taxonomy and Van den Hoven's (2007) moral reasons to protect privacy.....73

Table 3: Indicators to analyze consensus for tort laws.....81

Table 4: Indicators to analyze consensus for probability of occurrence and negative impact.....82

Illustration 8: Ratings on the expertise of the respondents.....86

Illustration 9: Probability and Impact scores for the 11 privacy incidents (materialized threats) that experts identified and their merger into three different groups of incidents.....90

Illustration 10: To what extent do the American tort laws address privacy threats in Social Network Sites and the resulting damages?.....93

Table 5: Framework applied to Beacon case.....101

Illustration 11: Data collection and analysis from a SNS and another source. Adapted from Aleman-Meza et al. (2006).....102

Table 6: Framework applied to Conflict of Interest Detection case.....104

Table 7: Framework applied to Megan Meier case.....106

Table 8: Framework applied to Samer Elatrash case.....108

Table 9: Framework applied to Jodie and Amanda Hudson case.....109

Table 10: Framework applied to Facebook Social Ads case.....110

Table 11: Framework applied to Facebook Third-party Applications case.....112

Table 12: Framework applied to Alex Hill case.....114

Table 13: Framework applied to Dutch 'pedophile' case.....115

Table 14: Framework applied to Hyves case.....116

Table 15: Framework applied to Doubleclick / MySpace case.....118

Table 16: The framework to deconstruct privacy incidents.....126

Illustration 12: Privacy incidents for users of Social Network Sites.....127

Preface and acknowledgements

This report is the final product of David Riphagen's Master Thesis for the Systems Engineering, Policy Analysis and Management (SEPAM) program at the faculty of Technology, Policy and Management (TPM) at Delft University of Technology. Other final products of the thesis are a scientific paper, a public presentation on this subject and a conference on October 23 and 24 about 'Privacy in Social Network Sites' at Delft University of Technology.

The idea of investigating privacy issues in Social Network Sites originated from my fascination for the interactions between information technology and society on the one hand and issues that we judge as immoral but are not illegal on the other hand. After writing a paper for the Delft University of Technology's Honours Track on 'Internet and privacy' and a paper about the topology of Social Network Sites, I discussed my proposal on Privacy in Social Network Sites with Milton Mueller, who just started as a visiting professor on 'Security and Privacy of Internet users' at my faculty.

Most of my research and data collection took place at the Electronic Privacy Information Center in Washington, DC, USA. But this Master Thesis is truly a product from all over the world. I have written contributions in Washington, DC, at my temporary home there, at Jennifer's home and at the office at EPIC. But also at the airport in New Haven, Yale University and the Computers, Freedom and Privacy conference in New Haven, during my short roundtrip of upper state New York in various Holiday Inn Hotels, during my short stay in New York City, at Delft University of Technology and at home in Delft, at the 'Data Retention in Europe' workshop and the hotel in Budapest, Hungary and even on transatlantic flights.

First, my thanks goes out to the members of my graduation committee, who have struck a difficult balance between supporting me in all my ambitions while at the same time keeping me with my feet on the ground and wisely advising me about my research. I want to thank Milton Mueller for helping me find the right resources at EPIC in the USA, Jan van den Berg for helping me with structure of my thesis, Jeroen van den Hoven for supporting me in organizing the conference and Scott Cunningham for specifically providing advise about my survey. Although not a member of my graduation committee, my friend Maarten Kroesen must also be mentioned here for helping with the data analysis of my survey.

Furthermore, I would like to thank the Electronic Privacy Information Center in Washington, DC, USA and in specific Marc Rotenberg, for being more than my host organization during my research. Furthermore I want to thank Melissa Ngo for helping me with my survey at EPIC, John Verdi from EPIC for the interesting and challenging debates during the long hours at the office, Lillie Coney from EPIC for learning me how to cold call people and helping me in perfecting my survey , Guilherme Roschke from EPIC for discussing his valuable insights on Social Network Sites with me and the EPIC clerks.

I am very thankful to Niels Huijbregts and Mieke van Heesewijk, for sponsoring my stay in the United States of America and the opportunity to post my weblogs on their website. The ability to post my ideas on such a significant platform really helped in getting my ideas across, not in the least place to the media.

Furthermore, I would like to thank the University Fund Delft, who helped me with funding for my travel costs to the United States of America. Without their contribution,

obtaining research results from the United States and working at the Electronic Privacy Information Center would have been much harder or impossible.

With respect to the media, I also want to thank Carinda Strangio from TROS Radio Online for asking me to give my opinion on the new features that Hyves implemented. This performance on prime-time was a good proofing for my ideas. Also, thanks to Marie-José Klaver from NRC Handelsblad and Hester Otter from Trouw for bringing my message across in the printed media.

I want to thank Christiaan Mooiweer from the weblog Geen-commentaar, who also provided a platform for my thoughts and posted some of my weblogs.

My great love and respect goes out to Jennifer Evans, who has provided me with moral support , even when I thought I would never be able to find a way out of all the data I collected and who also proofread most of my work. Not to mention that she was always there for me, in the United States of America and now here in the Netherlands.

I want to thank Eddan Katz from Yale University for the ability to present my viewpoints on Social Network Sites and third parties at the Computers, Freedom and Privacy 2008 conference in New Haven and letting me organize a Birds-of-a-Feather session on this subject.

Finally, special thanks to Jolien Ubacht, for being supportive on my initial ideas for a conference on 'Privacy in Social Network Sites' and for supporting me in this idea and in my dream to graduate in the USA.

Summary

Social Network Sites (SNS) are websites that allow users to upload information to a public profile, create a list of online friends, and browse the profiles of other users of the SNS. These websites have membership rules and community standards. Users disclose identity-relevant information via their profile to others. This information is either referential, directly referring to a person, or attributive, describing attributes to the data subject. Although most laws and regulations restrict the access to referential information, attributive information is not protected as such. However, the aggregation of large amounts of attributive information in SNS profiles poses new privacy risks.

Information spreads faster through a Social Network Site than through a real-life network. Information might be disclosed to a group of people unexpectedly, because the digital information is easily copied, can be stored indefinitely and is searchable. It especially harms users when information travels through different social spheres, and ends up with people whom it was not intended for.

Furthermore, Social Network Sites have financial incentives to generate revenues from the information users upload. Most of these websites are free for use, and SNS have to recoup the incurred costs by generating revenue from the identity-relevant information of their users. The most common method is to create marketing profiles of users and serve them with targeted ads. As SNS and their marketing partners obtain more information about their users, informational inequality arises. Because the SNS have more information about their users and users are not in a position to bargain about the terms at which they disclose their information, an architecture of control emerges, leaving the users vulnerable to harms. The constant surveillance of users, with or without the user's informed consent and mostly without the user knowing, makes Social Network Sites alike an online version of the Panopticon. This prison was designed by Jeremy Bentham, in such a way that guards could always observe prisoners, without the prisoners being able to see the guards.

Indeed, digital files of users are maintained that attempt to resemble the real person as closely as possible. Other users can contribute to this profile by uploading photos or text about that user, often without his informed consent. This raises serious questions about the user's control over his identity-relevant information and his ability to construct his own moral identity within a SNS. Social Network Sites also collect information from other websites. When this happens without his consent, the user is prohibited from creating his own moral identity.

A fourth reason to restrict access to the information is the prevention of information-based harm. Some activities that harm users need specific information on the data subject before they can be executed, such as a Social Security Number to obtain credit. Indeed, your address and current employer can be used to determine when you are not home and break into your house.

The activities that harm users are grouped into information collection, information processing and information dissemination. This classification helps in identifying the specific activities that cause harm to users, and supports designing measures to prevent these activities and the damage they cause.

The survey shows that there are three main privacy risks for users of Social Network

Sites: Total Information Awareness, dissemination to wrongdoers and no control over your identity-relevant information.

Social Network Sites track the activity of their users on their own websites and those of their marketing partners. They are able to gather unprecedented amounts of secondary personal information on their users, sometimes even without the informed consent of the users. An architecture of vulnerability emerges when the government collects this information and uses it to control its citizens. A Dutch judge found the private profile of a SNS user public, because people can be added to the friends list easily. Studies for the United States government have shown that it is easy to collect data from public Internet sources as SNS and connect them to existing government databases. This could lead to the chilling of free speech with respect to political dissidents.

Because of the great amount of identity-relevant information, which disseminates easily through a Social Network Sites, this could end up easily with wrongdoers. Stalkers, bullies and predators use the attributive information on SNS to identify with their victim and use the referential data to contact them. The profiles of users combined with the ease of contacting a user make SNS a useful platform for wrongdoers. The information on the websites can also easily be used to damage someone's reputation, and with the large amount of attributive data on SNS, it is not difficult to reverse engineer information needed to steal someone's identity. Although there is no proof that these things are affecting all users of SNS, experts agree that they affect a significant amount of users and can cause great damage to the users.

Social Network Sites interpret the consent that users give when signing up for their services as a broad informed consent, which can be stretched to secondary usage. In reality, users have minimal information and no control over secondary use of their information, the selling of their information or the disclosure of their information to groups unwanted, by the SNS. Above all, others can post information about the user, which can only be deleted after the fact, if possible at all. Information is posted about non-users, but they cannot delete this, unless they become members.

Conventional laws and regulations do not address these issues. Of the American tort laws, only the publication of private facts tort and the appropriation tort seem to address the problems mentioned above. However, it is hard to proof that the facts are private when a user posts them on a SNS profile and the monetary damage is in both cases difficult to measure. Social Network Sites violate many of the Fair Information Practices as set forth by the OECD and recognized by many countries. The use of the information is not limited to the specified purpose and the processing of the information is very covert. The privacy watchdogs in various countries do not have the right means to sanction SNS for violating the Fair Information Practices.

A more colloquial approach is needed. Harmful activities should be grouped into information collection, information processing and information dissemination and harm should be defined by the four moral reasons to restrict access to identity-relevant information: information-based harm, informational inequality, information injustice and the inability to define one's own moral identity. The activities that cause specific harms can be identified and constrained. It is recommended to design and start a policy development process in which relevant actors jointly identify preventive measures to minimize privacy risks for users of Social Network Sites.

1. Social Network Sites: increasing in numbers and problems?

1.1. Problem Analysis

Social Network Sites[1] (SNS) are growing by any metric. The numbers vary, but the common denominator is growth: whether in the amount of SNS worldwide (Kumar et al. 2006), the number of users (Schonfeld 2007), the monetary valuation of SNS (Gross 2006) and the amount of identity-relevant information that is stored in these Social Network Sites (Gross et al. 2005). Their growing and enduring appearance has attracted the attention of the media, researchers and privacy activists. The European research agency ENISA (1997) mentions that "[Social Network Sites] are one of the most remarkable technological phenomena of the 21st century". Users of SNS create profiles that resemble their personality and disclose personal information as name, interests and whereabouts to members of the websites (and sometimes beyond). This could lead to the unwanted retention or disclosure of information. In particular, the unregulated retention of profile data by SNS Facebook has gained the attention of the British Information Commissioner's Office, which has started an investigation (Vallance 2008).

The user of a Social Network Site is both creator of the profile and consumer of the information on others' profiles. This has lead to outcries among users (Noguchi 2006) when SNS implement new features that disclose previously unlisted personal information to a wider audience than anticipated by users (TechNews 2006), which has potentially serious privacy implications. Barnes (2006) calls this the privacy paradox[2]. She mentions that "[y]oung people are pouring their minds, if not their hearts, into cyberspace. They are doing it to clear their heads, stow their thoughts and get feedback from peers." However, these users assume a certain amount of control over who sees this disclosure. Users expect that they can fully determine to whom they disclose their information. The observation that "[t]eenagers will freely give up personal information to join social networks on the Internet, [...] [but afterwards] they are surprised that their parents read their journals" (Barnes 2006) goes to the heart of the privacy paradox.

It is this fundamental problem that illustrates the complexities surrounding Social Network Sites. A common trade-off in informational privacy, according to Van den Hoven (2007), is that "[m]any customers have no problems with alerts [from] businesses [that] draw their attention to bargains in the areas they are interested in [..]". In other words, people are willing to give others access to their personal information if they think they derive benefits from it. The choice to disclose identity-relevant information[3] is based on the utility function of the user that balances the gains of disclosing information with the costs of possible harms. SNS fully embrace this behavior and ultimately hope to profit from collecting, processing and disseminating their users' personal information.

The abundance of easily accessible identity-relevant information on SNS can be used for

---

[1] Refer to paragraph 2.1 and the glossary in chapter I on why I use the term 'Social Network Sites' instead of 'Social Networking Sites'.

[2] Refer to the glossary for a broader explanation.

[3] Refer to paragraph 2.3 and the glossary for a definition of the term 'Identity Relevant Information'

harmful activites In this respect, Social Network Sites can be regarded as the online version of the Panopticon prison. In the eighteenth century, Jeremy Bentham developed this prison, in which the prisoners could always be observed by the guards, without the guards being visible to the prisoners. This metaphor has been used by Foucault to describe societal surveillance (Lyon 1993). Social Network Sites are able to collect information about their users from diverse sources, without the user noticing this. While the Social Network Sites can surveil its users, the user is uncertain of when this snooping takes place, thus resembling a online panopticon.

Owners of SNS have made several attempts to benefit from the personal information stored on their servers, and this has lead to public outcries and scrutiny from activists, researchers and regulators. Many of whom believe these activities are illegal or unethical. A good example is the introduction and integration of third parties that directly benefit from the profile information and connections of SNS (Soghoian 2008). It is one thing to make this observation, but it's quite another (1) to understand precisely why and how SNS try to profit from users' information, and (2) why users are willing to give away so much details about their lives.

1.2. Previous research

Previous research on privacy issues in SNS focuses either on identity management, reputation, moral issues, information flow or is the result of research by public interest groups or research agencies.

danah boyd has published several papers on SNS as Friendster (2004), MySpace (2006) and Facebook (2006a). Her work focuses on the performance of creating an online identity, especially by teens, on the different SNS and what the reasons are why SNS are so popular with young people (boyd and Heer 2006b). She has structured the characteristics of SNS and provided an analysis of whether SNS should be considered public or private (2007a).

Daniel Solove takes a more legal approach to issues caused by (new) information technology, and has created a privacy taxonomy (2006) to identify specific activities that pose threats to privacy. He distinguished between information collection, information processing, information dissemination and invasions. One of his books, 'The future of reputation: Gossip, rumor, and privacy on the Internet" (2007a) deals specifically with the privacy issues on Social Network Sites. Solove describes the risks for reputation that weblogs and SNS pose to their users.

Jeroen van den Hoven and Noëmi Manders-Huits (2006) focus on the definition of the information that we want to keep private or restrict access to and why we want restrict access to this information (Van den Hoven 1997 and 2007). There analysis includes various types of information technologies, but is not specifically directed at Social Network Sites.

Gross and Acquisti (2005) and Acquisti and Gross (2006) have conducted an empirical study on what types of information users post on their Facebook profile, how easy it is to access this information and what privacy risks are the consequence of this. Felt (2008b) looks at Facebook as well, and specifically at the information flows between the SNS and third-party applications. She finds that the access that these applications have to user information poses

serious risks for privacy.

Various public interest groups, as the International Working Party on Data Protection in Telecommunications (2008), the Canadian Internet Policy and Public Interest Clinic (Lawson 2008), the Electronic Privacy Information Center (EPIC 2008) and the European Network and Intelligence Security Agency (2007) have published reports or websites about Social Network Sites, the threats they pose to privacy and recommendations to prevent these threats from occurring. The EPIC website and CIPPIC's complain with the Canadian Privacy Commissioner describe the specific incidents that threaten privacy very well, but lack the generalization needed to provide policy makers with recommendations. The IWPDPT's report and ENISA's position paper are in-depth studies based on expert consultation. Both papers present recommendations to prevent privacy harms for users of SNS, but they do not present their methodology for identifying or prioritizing the risks. Nonetheless, both documents are very valuable for expert consultation.

A holistic study of SNS and the privacy risks for their users has not been published to this date. Furthermore, agreement between experts on which privacy risks need to be prevented immediately, is non-existent. With this research paper, I will provide more insight in the different trade-offs that users face and the interactions between users, SNS and third parties. In addition, I will provide a survey of the most pressing problems for privacy as identified by privacy experts.

1.3. What this study achieves

The goal of this study is to create better insight in the complexities of privacy in Social Network Sites. When the various trade-offs and relations become more visible, policy makers can assess the vulnerabilities and come up with preventive measures (Van den Berg 2008). This thesis must be seen in the light of a larger process of policy formulation. The rational model of policy formulation as described by Bots and Hulshof (1995) based on Simon (1997), which is shown in illustration 1, shows how a policy formulation takes place. Although this is a significant simplification of the real policy process, it clearly depicts the different steps in policy formulation and the positioning of this paper. This is important, because it allows a separation of concerns when dealing with complex problems like privacy risks for users of SNS. It is very difficult to come up with good policies if the problem is not sufficiently structured. Without knowing the problem thoroughly, preventive measures could address the wrong issues or even increase the vulnerabilities by other threats.

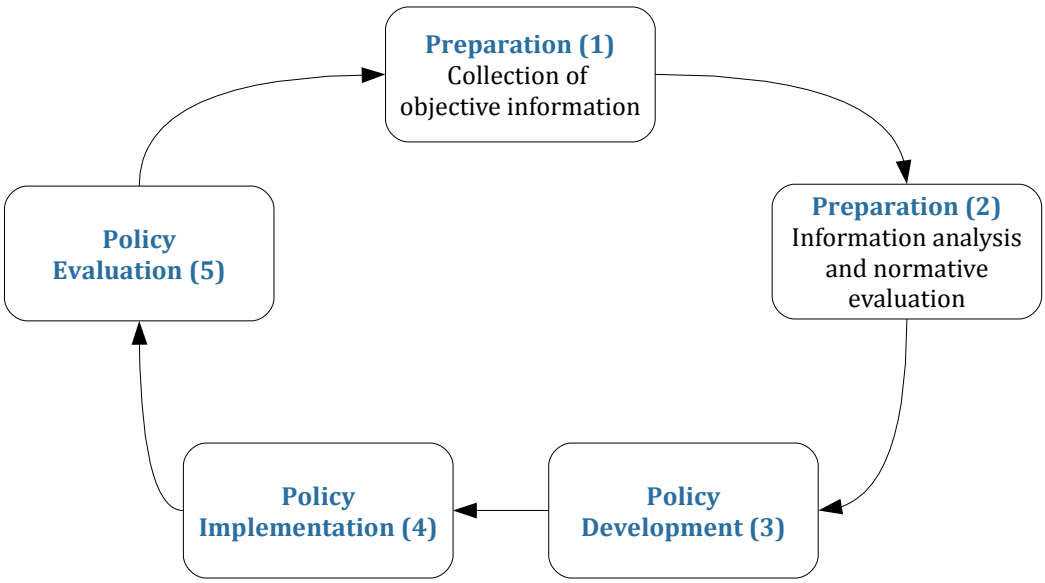


Illustration 1: A rational model of policy formulation, adapted from Bots and Hulshof (1995)

This thesis focuses on the preparation of the policy process by identifying the privacy risks for users of SNS. The model, shown in illustration 1, distinguishes between the collection of objective information (1) and the analysis of this information and providing a normative evaluation (2) (Bots and Hulshof 1995). The first step is part of the exploratory research, or collecting and structuring all the necessary information to provide a usable overview of the problem. Only when there is a good and elaborate overview of the problem does it make sense to go to the second step to assess the situation. The second step consists of constructive research, or the construction of a model to analyze the situation and provide a normative evaluation. With this evaluation, policy makers are provided with a prioritization of issues to address in their policy development. After the policy is developed, it can be implemented and evaluated. This rational model depicts the policy process on a very high level, and does not take into account the different perceptions and often conflicting interests of actors, pluralism, dynamics and interdependence. These aspect will be described in chapters 4 and 5 of the exploratory research phase.

1.4. Demarcation

This paper is part of my graduation research project, and because there is a limited timeframe for this project, not all issues concerning privacy in Social Network Sites can be addressed. Therefore I propose the following demarcation:

- The geographic area of the research is initially constrained to SNS that are owned by USA-based private companies. The reasons for this are that most of the SNS are American-owned, American SNS have the largest user base and more of the American SNS are being scrutinized by regulators, researchers and privacy activists in the United States (such as the Electronic Privacy Information Center and the Electronic Frontier Foundation).

- By focusing on SNS owned by private organizations, I stay out of the complex field of balancing national security with privacy concerns, and government interests versus

citizen's interest, which are whole topic in themselves (Solove 2006).

- Because of time constraints and the observation that users of SNS seem to underestimate the privacy threats in SNS (Jones and Soltren 2005, Gross 2005), I did not conduct surveys among users of SNS with the goal to identify privacy harms.

In this study the emphasis is on identifying privacy risks for users of Social Network Sites. Such an emphasis does not deny, discount or diminish the social, educational an economic value of SNS (ENISA 2007). I argue that by identifying the privacy risks and harms and providing policy recommendations to address these harms and risks, Social Network Sites can become a safer place and users will be able to reap the benefits from information sharing, without having the drawbacks that they are now subject to.

1.5. How this study achieves this

The main research question of this paper is:

What are privacy risks for users of Social Network Sites?

Sub-questions include:

1. How to define and characterize Social Network Sites, identity-relevant information in SNS and explain why the access to this information should be restricted?

2. What social (law, regulatory), technical (topology) and economic (value chain, flow of financial means) characteristics describe Social Network Sites and how do these influence the trade-off between privacy of users and self-disclosure? How can these characteristics be structured?

3. How can the trade-off between restriction of access to identity-relevant information and the will to disclose this information via SNS be deconstructed and how does this deconstruction provide insight in possible privacy harms?

4. What privacy incidents do users of SNS face, what is the prioritization of these incidents and how do the threats that cause these harms interact with each other?

5. What are examples of privacy incidents and how do these fit in the developed taxonomy of privacy in SNS?

6. How to validate the framework for privacy threats, incidents and harms?

7. What considerations should policy makers take into account and how can they use the framework to come up with effective policy measures?

8. What would a research agenda devoted to a better understanding of the privacy risks for users of SNS look like?

Illustration 2 shows how these sub-questions are treated within the different chapters of

the thesis, which theories (methods) I have used to answer these question and which activities (techniques) I have used to gather and analyze data. Chapter one is an introduction and uses the rational policy model of Bots and Hulshof (1995) to show how this study relates to other activities in the (rational) policy formulation process. In chapter two I will define the most important concepts for the study. As mentioned in paragraph 1.2, Daniel Solove (2006) provides an extensive overview of the various activities that pose threats to privacy. He has applied his privacy taxonomy to informational privacy issues, of which privacy issues in SNS are a subset. Van den Hoven (2007) provides moral reasons to restrict the access to identity-relevant information, and shows where the exact harm from privacy incidents comes from. However, its is important to distinguish between threats for privacy that never occur and threats that do materialize, and identify the likelihood of them occurring. While Van den Hoven's work (2007) shows how damaging the specific harms are, risk analysis as proposed by Van den Berg (2008) provides a good framework to distinguish between threats that occur and the vulnerabilities that enables them. This form of risk analysis also clears the way for a prioritization of privacy incidents, which helps policy makers with identifying which problems they need to address first. To provide a complete overview of the key features of SNS, I used the MCM framework of Hoegg et al. (2006). This framework was specifically developed to overcome the shortcomings that regular frameworks have in describing new Internet-based businesses. I will use this framework for the exploratory research in chapter 4. The OECD's Fair Information Practices are world-wide seen as the standards for data collection, processing and dissemination (Gellman 2008). In several countries government agencies have to abide by these Fair Information Practices and they are a good reference for judging the data processing practices of Social Network Sites. There are several privacy threats for users of SNS and some of these threats are made up out of different activities that, when executed together, harm the user. In chapter 5 I will construct a framework to analyze these threats and determine exactly what different activities add to the privacy threat. As mentioned before, the work of Van den Hoven (2007) and Solove (2006) is used to define this framework. This framework is presented in chapter 5, which is still part of the exploratory research, as it constructs the framework by which to analyze the results from the survey.

The third part of this thesis consists of empirical research. The results from the previous chapters are tested with empirical data and a survey is used as a mirror for the framework developed in chapter 5. The methodology of the survey is described in chapter 6 and I draw upon the work of Bots and Hulshof (1995) to define disagreement among the experts that participated in the survey. They recognize in their work that the measurement of disagreement among a group of respondents is important to validate the results. These results are presented in chapter 7, where I will analyze them with the help of a probability-impact matrix. Although I do not use probability in a mathematical way here, the best way to describe the ratings that the experts performed is 'probability of occurrence on a large scale'. In chapter 8 I will use the privacy incidents as identified by the experts and analyze them with the framework I developed in chapter 5. The eleven incidents that the experts identified, will be deconstructed with the framework. In this regard, the empirical data provides a test case for validating the framework. In chapter 9, I will reflect on the methodology used and the substantive results of the research. Also, the options for implementing the outcomes of this thesis in policy recommendations is assessed, and a process for policy formulation proposed. Chapter 10 presents conclusions and recommendations for future research and policy makers.

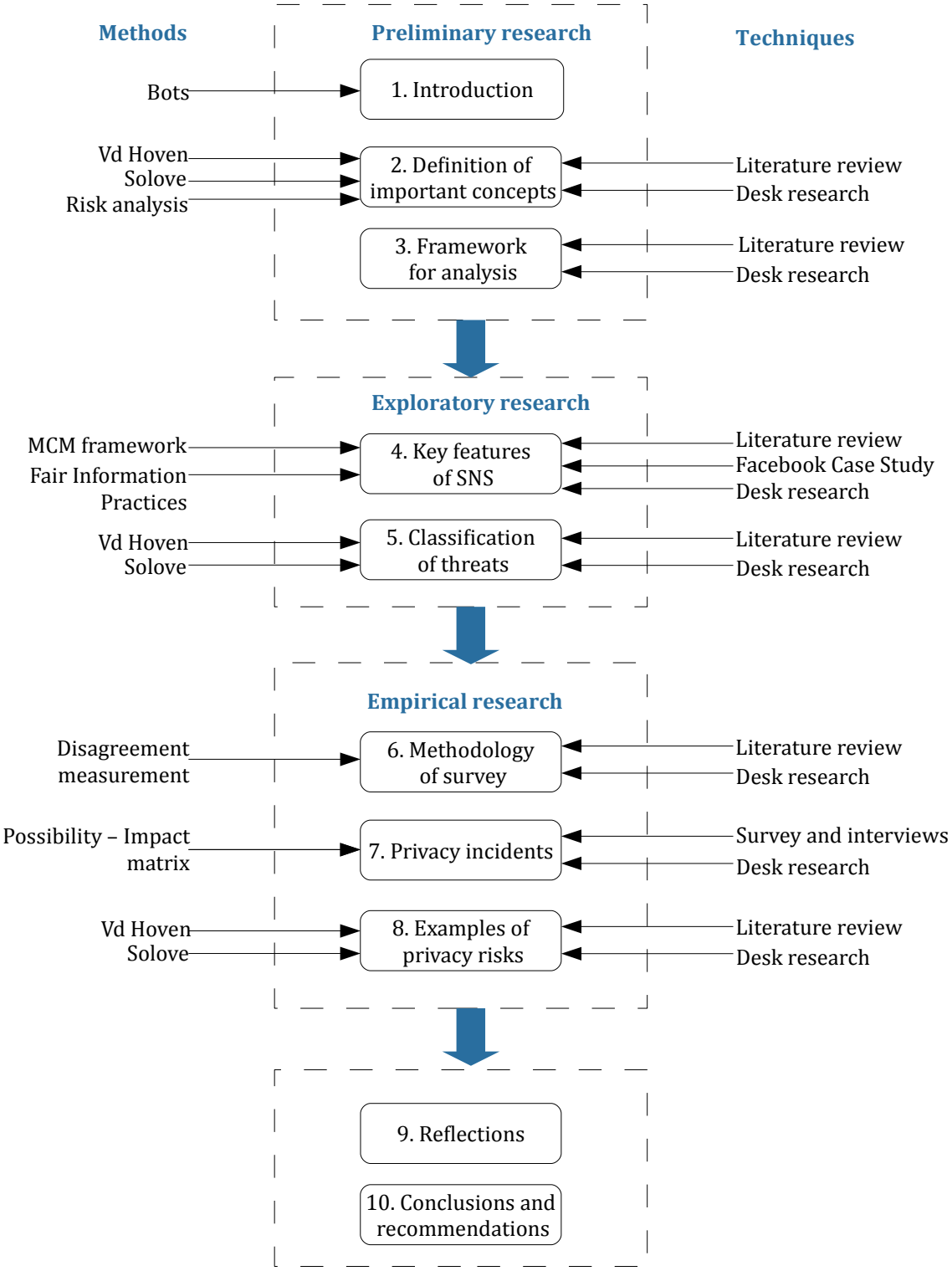


Illustration 2: Research approach

I have addressed these questions in my research, conducted at the Electronic Privacy Information Center (EPIC) in Washington, DC, USA. In illustration 2 the various techniques used during the research are also depicted:

1. A literature review. More than hundred sources on various topics, ranging from books to scientific articles to newspaper articles and documentaries, have been surveyed and included in this research. I have reviewed the works of Jeroen van den Hoven, Daniel Solove, Marc Rotenberg and danah boyd extensively. Please refer to appendix II for a complete list of references.

2. Desk research. I have reviewed a wide breadth of literature and it is not always evident how this literature relates to each other and to the problem at hand. Therefore I conducted desk research, while in Washington, DC, and consulted frequently with my co-workers at EPIC. This is how I assembled the framework used in chapter 3 and how I created the framework to analyze privacy incidents in chapter 5. Desk research also included a thorough analysis of the data I obtained from the survey and the interviews, as mentioned in chapter 6.

3. A case study of an American Social Network Site, Facebook. Facebook is the second-biggest SNS in the USA, after MySpace (Nielsen 2008) and has been the center of attention because of issues surrounding privacy multiple times (EPIC 2008). The main goal of the case study is to evaluate how the Facebook policies, default settings and privacy controls create a fair playing field for both Facebook and its users. Do the users have bargaining power, can they unbundle services or are they locked in to potentially harmful outcomes?

4. A Survey among American privacy and Internet experts. The field of research on privacy in SNS is scattered: it ranges from moral analysis, to behavioral science, law and graph and network theory. To integrate the various viewpoints on SNS and disciplines, I have conducted a survey among more than 30 American privacy experts. I was able to survey and interview these privacy experts because of my relationship and hosting with the Electronic Privacy Information Center. I have included a half dozen of the interviews I conducted in appendix B.

For readers interested in how this study answered the questions above, chapter three describes the encompassing framework to study privacy risks in SNS. Chapter five describes how the various privacy threats for users lead to incidents and how we can assess whether these incidents are harmful or not. And in chapter six, I describe how I acquired input from various privacy and Internet experts and asked them to identity incidents that lead to privacy harms.

A glossary can be found in appendix I and in appendix II you can find the references.

2. What are Social Network Sites and why can they harm us?

2.1. What is a Social Network Site?

The Internet is a network of networks that consists of computers that are all connected to each other. Transport and addressing of the packets that travel over these networks is defined by the TCP / IP protocols. These open protocols define the Internet as consisting of four communication layers (1989), also known as the TCP / IP stack:

1. The application layer,
2. The transport layer,
3. The Internet layer and
4. The link layer.

The layers operate independent from each other and each layer builds on the functionalities of the lower layer. For example, the link layer provides the physical connections for the Internet layer to route packets over. The transportation layer makes for the reliability of the transmissions and this network of hard- and software provides the backbone for applications such as FTP, bitTorrent, SMTP, POP and the protocols that make up the World Wide Web (HTTP, HTML). Social Network Sites make use of these applications to provide their functionalities. It is important to remember that the World Wide Web (WWW) and the Internet are not the same and thus have different topologies. A server might well function as an e-mail server (POP or IMAP) only, without being part of the WWW. It could, for example, not host any websites.

Social Network Sites use the latest technologies on the highest layers of the Internet as defined by the TCP / IP or the OSI model. While all SNS are based on databases containing profiles, algorithms used to collect, process and disseminate information and the code that provides a Graphical User Interface (GUI) are different per website. Facebook uses proprietary algorithms to determine which information from your profile is disseminated to which friends[4]. Their algorithms are even patented (US Patent & Trademark Office 2007).

The GUI that SNS provide is aimed at making it easier for users to upload personal information. It is designed to cut away contextual cues regarding to whom the information might be disseminated. In an interview with professor Solove (2008), he expressed his concerns about the omission of contextual cues in SNS on disseminating information to unwanted parties. In another interview, Verdi (2008) mentioned that

> *"The perception that SNS cultivate doesn't match with the reality. They cultivate the view that you can submit information to a SNS and control where the information goes. There is a disjunction between that perception and reality. In the reality this information is available to many individuals, many organizations and many entities in a variety of ways in a very uncontrolled, unmonitored and unaudited way."*

[4] The term 'friends' will be used to identify all the people that users have in their contact list. Although they technically do not have to be friends in real life, most SNS refer to them as friends. Contact list would also be an appropriate name, which refers more directly to contacts as used in SNS LinkedIn, but researchers have stuck to the terms 'friends' and 'friends list'.

Social Network Sites are Internet services based on databases of profiles and they provide a GUI for users to upload information and download information from other profiles. The user is both consumer and producer. Ellison and boyd (2007) state that SNS are part of online mediated environments. Mediating technologies, such as television, radio, newsprint or websites, change the scale of the public (boyd 2007). With other words, messages disseminated via a mediated technology can reach a larger audience. Mediated technologies, such as newspapers and televisions, have been around for more than an century. But when mediated technologies were connected to each others via networks, the messages in these environments could be disclosed to even a larger audience. In the networked environment, it has also become possible to search for messages. The online (networked) mediated environments have the following characteristics: persistence (what you say sticks around), searchability of the information, replicability (digital bits are easy to copy) and invisible audiences (you are not able to fully determine who you are communicating with) (boyd 2007). This means that information you would rather not have online stays online forever, everyone is in potential able to find it via search functions and can then copy it to its own computer or other websites. It is not hard to imagine situations in which this is undesirable

Ellison and boyd coined the term Social Network Sites, which is more appropriate than Social NetworkING Sites. They argue that the emphasis of SNS is to articulate and show your social network and not so much to connect to new people (networking). Furthermore, I add that SNS derive much of their value from making use of the connections between you and your friends (your network), such as disseminating status updates of your profiles to friends (Facebook's Newsfeed), showing you what your friends bought (Facebook's Beacon) or letting you know what applications from third parties your friends have added to their profile (Facebook's Application Platform)[5].

As mentioned before, boyd is an authority on describing SNS and the trade-offs users make between privacy and social interaction when using these services. Building on Ellison and boyd (2007), the position paper of the European research agency ENISA (2007) and my earlier paper on the topology of SNS (Riphagen 2008a), I propose the following definition of Social Network Sites:

Social Network Sites are websites that:

1. Primarily allow users to divulge information (self-disclosure) by constructing a public or semi-public profile (ENISA 2007), which contains identity-relevant information (referential as well as attributive), within a bounded system, and which information is stored by the SNS in databases.

2. Enable users to articulate a list of other users (friends list)[6], with whom they share a connection, and make this list visible to other users to facilitate contact between users and interact in various ways with these users (ENISA 2007), sometimes by making use of the proprietary algorithms that SNS provide.

3. Enable users to view and traverse their list of connections and those made by

[5] For a more elaborate view of Facebook's usage of your social network, see the Facebook case study in Appendix A.

[6] Refer to footnote 4 on SNS 'friends' and friends' list.

others within the Social Network Site, by means of a graphical user interface on top of the database of profiles, and enable users to: connect to one another with bilateral agreement; define the social relationships they have with their friends (ENISA 2007) leave public testimonials or comments (boyd 2007); or create and join communities that share the same interests and activities (Fu et al 2007).

4. Are characterized by membership rules with their own community standards and sustained social interaction between their members (Rothaermel 2001 based on Lawrence 1995 and Karp et al. 1977).

Examples of Social Network Sites are Hyves (Dutch), Facebook, MySpace, LinkedIn (for business professionals), Sugababes (Dutch) and Bebo. They can be grouped by whether they are activity-centered, identity-driven or affiliation-focused (Ellison and boyd 2007). An example of activity-centered SNS is couchsurfing.com, a SNS that connects people who want a free place to sleep while traveling (couchsurfing) to people offering an empty couch or spare bedroom. The various dating websites are also examples of activity-driven SNS. Other activity-driven SNS that focus on hobbies are websites as YouTube and Flickr, which incorporate the functions of SNS in their websites. Identity-driven SNS are, for example, BlackPlanet, directed at black people or Classmates.com, which aims to bring former classmates in contact with each other. Affiliation-focused websites are less common, but MyChurch is a good example of a SNS for people with an affiliation to a church. Recently, the campaign of American Democratic presidential candidate, Barack Obama, endorsed an affiliation-focused SNS by adding MyBarackObama to their website. Co-founder of Facebook, Marc Hughes, has taken on the challenge to gather supporters of Obama on their SNS and help them create events, messages and fundraising activities (Stelter 2008).

2.2. What is identity relevant information and why is it important?

Users of SNS upload various types of information: details about their whereabouts, pictures, videos, blogs and personal information such as name, phone number and address. The European Union has implemented a directive to protect individuals with regard to the processing of personal data and the free movement of personal data (2005). The definition that the EU gives of personal data is:

"(a) 'personal data' shall mean 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;"

This is a broader definition than 'Personal Identifiable Information' (PII), which constitutes only the information that directly and uniquely identifies a person and is used by TRUSTe, the Internet industry's own privacy seal program (Benassi 1999). The problem with PII is that different pieces of information that cannot identify you uniquely, can be collected and aggregated in such a way that they do identify you. A commonly used example is the American Social Security Number (SSN). This nine-digit account number was brought into being to administer the USA's Social Security laws (EPIC 2008a). The number consists of three numbers keyed to the state in which the number was issued, two numbers indicating the order in which the SSN was issued in the state and four randomly generated numbers. While the state number and the order number are not PII, they can be used to reverse engineer the

SSN and uniquely identify a person. Identity thieves can buy SSN from black markets and by trial-and-error pair the right numbers with the right persons to get credit, as Internet security firm McAfee found out (Paget 2008). Furthermore, as many credit agencies do not match names and SSN, getting credit by using another name than is coupled with the SSN is common practice (EPIC 2008a). Note that the state where the SSN was issued and the date of issuing would be 'personal data' under the EU definition, as they can be used to identify a person by referring to these factors.

An additional concern lies in the fact that although some information is not useful for identification purposes in its own right, it can be (indirectly) harmful and is thus considered as sensitive. People prefer that this information to remains private. For example, disclosure of one's income could make one more prone to crime and one may wish this does not become publicly available.

The potential for harm underscores the importance of defining and protecting personal information. Manders-Huits and Van den Hoven (2006) coined a definition of personal information that better suits the Social Network Site's environment: 'identity-relevant information'[7]. Van den Hoven (2007) mentions that data protection equals the restriction of access to personal data. He states that rough data can be transferred into meaningful information or knowledge, and because this could harm us, we should restrict access to such information. According to Manders-Huits and van den Hoven (2006), the definition of the European Union is too narrow, because if two facts about a person that are not initially considered personal information are combined, that person could well be identified and harmed. Van den Hoven (2007) gives the example of a priest telling the exceptional story of his first confessor (not personal information according to the EU), who confessed to a murder. Later in the conversation, a man walks into the room. When this man is asked how he knows the priest, he mentioned that he was his first confessor, which is not personal information, but still harms the man (Manders-Huits and Van den Hoven 2006 based on Gavison 1980).

Manders-Huits and Van den Hoven (2006) distinguish between referential data, that refers directly to one person and identifies him; and attributive data, that describes the characteristics of a person, or what is attributive to him. An example of referential data is the SSN, and an example of attributive data is hair color. This is a useful distinction, because aggregation of attributive data, as happens in the profiles of SNS, could identify and harm a person. Specifically, the whereabouts and preferences that users of SNS post on their profiles can be used to uniquely identify them. The profile as a whole is referential and this reference can be used to facilitate what is commonly known as price discrimination. Price discrimination[8] is charging different prices for the same good of service to different groups of consumers.

---

[7] This term was first coined "Identity Related Data" in 2006 and later translated by van den Hoven into "Identity Relevant Data" in 2007, which is the term that I will use in my thesis.

[8] The common perception is that profiling and price discrimination lead to lower prices and bargains for people who are interested in specific products. However, the legal areas that surrounds price discrimination are dark-grey. Privacy scholars have argued that price discrimination by means of profiling may make all consumers worse off (Rotenberg 2001). Because privacy strengthens the user's bargaining power, profiling makes his bargaining power less. As Rotenberg (2001) mentions "[w]hen you combine market power, consumer profiling and price discrimination, consumers may be less well off."

Both types of information are grouped under the header of "identity[-]relevant information" (Manders-Huits and Van den Hoven 2006, Van den Hoven 2007), which is both broader than the more commonly used personal identifiable information and the legal definition of personal information used by the European Union. Their concern with storing identity-relevant information focuses on the fact that people are judged on the representation that is created from identity-relevant information in databases. This concern is also applicable to the profiles in Social Network Sites. Also, they mention that the protection of identity-relevant information is important to reduce obstacles for 'experiments of living'. It seems highly unlikely that one would want the information on their SNS profile disclosed when applying for a job (Manders-Huits and Van den Hoven 2006). I will use the term identity-relevant information to refer to the attributive and referential data that users upload to Social Network Sites and could provide harm to them, because this information is relevant to their identity. But why should we protect this information?

2.3. Reasons to protect users' information on Social Network Sites

In the previous paragraph I provided a definition of personal information that focuses on information that we would rather keep private, because it could identify us without our consent or because public availability of the information would make us more vulnerable to harmful activities. This quite simple definition stands in stark contrast with the broad and ongoing debate about privacy (Van den Hoven 2007). In the case of Social Network Sites and privacy, the focus on specific harms for users of SNS through the use of their identity-relevant information is adequate and productive.

In the kick-off meeting for my research project, Van den Hoven mentioned that a focus on data protection would be more appropriate because the discussion about privacy has been obfuscating (Van den Hoven et al. 2008). In addition, Mueller pointed out that we should not lose sight of the social implications for users of SNS (Van den Hoven et al. 2008). Focusing on identity-relevant information that could potentially harm us both incorporates Van den Hoven's view of protection of personal data and Mueller's point that both benefits and harms from using SNS derive from social interactions and network effects.

Problems with the concept of privacy include the dichotomy between private and public spheres, also known as the secrecy paradigm. Solove specifically addresses this secrecy paradigm, and agrees that privacy is a concept in disarray (Solove 2006). The secrecy paradigm is a guideline to identify privacy harms based on the concept of private or secret information[9]. If information is available to the public, it is not deemed private and therefore the privacy paradigm concludes that usage of the information is not harmful in terms of privacy. The secrecy paradigm focuses on breached confidentiality, harmed reputation and unwanted publicity. These harms are, according to Solove, not the central problems of databases (Solove 2004). The focus on privacy harms through the use of identity-relevant information does not have these complications. Public information, such as photos posted online and a real name posted on another web forum, can be combined and used to harm people. The secrecy paradigm does not recognize a harm here, while the focus on identity-relevant information and potential harms does.

Solove (2006) has drawn up a framework that classifies and describes the privacy harms

[9] Refer to the glossary in appendix I for a more elaborate definition.

in a broader spectrum. This privacy taxonomy is the only classification of activities that create risks to privacy based on a legal system and incorporating the new type of activities enabled by Internet-based services. Therefore, Solove's taxonomy (2006) is relevant to identify activities on SNS that could pose privacy risks to users of these websites. His analysis is law-oriented, and I find that the moral reasons to protect privacy that Van den Hoven (2007) provide a necessary complement to his framework. Solove describes harmful activities or threats. He distinguishes the following four activities and threats:

- Information collection,
- Information processing,
- Information dissemination, and
- Invasions.

Van den Hoven focuses more on the outcomes of these activities. What is the actual harm that can be incurred when a harmful mechanism is applied? He distinguishes the following harms:

- Information-based harm,
- Information inequality,
- Informational injustice, and
- Restricting moral autonomy and preempting one's ability to create his own moral identity.

The harmful mechanisms of Solove are less meaningful if we do not know why and how these activities could harm us. Similarly, the moral reasons for protection given by Van den Hoven have less explanatory power if we can't determine which activities lead to these outcomes. A combination is more powerful: in this case, the sum of one and one is three. In chapter five, I will describe Solove's privacy taxonomy in more detail and I will classify concrete examples from the analysis of SNS and match them with Van den Hoven's reasons why these examples are harmful.

Why should we protect people's identity-relevant information and constrain access to this information? In different papers, Manders-Huits and Van den Hoven provide a good overview of reasons we should constrain the access to this information. Most of their reasons can not be derived from legal systems, because harm is difficult to prove or not measurable in monetary terms. Their moral analysis, focusing on what identity is and why it is a valuable asset to humans, is unique in its broadness and in-depth analysis of moral issues concerning identity and its protection. The first reason they mention to constrain the collection, processing and dissemination of identity-relevant information is information-based harm. This is harm from crimes that are made possible or are more easily committed because of the collection of information. Identity theft is a good example. Refer back to paragraph 2.2 on how a Social Security Number can be obtained from a Social Network Site and used to hijack someone's identity. Accumulative information-based harm also falls under this header: releasing or collecting snippets of identity-relevant information at different occasions that together could be used to inflict harm upon a person. I refer to the example of connecting

photographs with first and last name mentioned in paragraph 2.2. This practice is common on SNS: it is called 'tagging' on Facebook and 'gespot' on the Dutch SNS, Hyves[10]. Users upload pictures of other people and tag them with the name of that person. Needless to say, because the subject in the photo does not consent to the placement of the photo or the attachment of his name to it, this could lead to harmful situations. An example that also incorporates informational injustice is the placement and tagging of a disgraceful party photo of a person by a third person and the disclosure of this picture to a potential employer because he conducts a background check.

This leads to the second moral reason to protect identity-relevant information: informational inequality. There is a market for identity information, and companies are collecting information about purchases and preferences to create profiles about consumers, which are useful for marketing purposes or price discrimination. Refer to footnote seven for the discussion on upward as well as downward price discrimination. Because most people are unaware of the market for their personal information, or because they are not in a position to bargain or unbundle the services they get in exchange for their identity-relevant information, there is an informational inequality. When a person signs up for a Social Network Site, that individual user does not have the bargaining power to determine how the information he uploads to his profile will be used. Although Europe has regulations about secondary use, users cannot bargain with a SNS regarding the use of their information. Of course, users do not have to fill out every field when they sign up for an SNS, but SNS are constructed in such a way that users are stimulated to upload as much information as possible. It is also impossible or very hard to unbundle the services that are being offered by a SNS. It is impossible for users of both Hyves and Facebook to opt-out of the photo tagging service described above. Facebook notifies tagged users and provides an option for 'untagging' the photo, but this is all after the fact.

Users are often not aware of the size of the audience accessing their content (ENISA 2007). Dissemination of identity-relevant information may then lead to harm, if the information ends up with an undesirable recipient. Consider again the tagged photo at the disgraceful party. What if the subject on the picture has no problems with his friends seeing the picture, but does not want his professors from the university to have access to the photo? Van den Hoven thinks this is a reasonable expectation and refers to Michael Walzers spheres of justice (Walzer 1983) to identify different social spheres of access. In these spheres we, as moral persons, want different kinds of identity-relevant information about us to be known. We do not want that our professors know about our parties or that our prospective employers know about our socially-unacceptable behavior. In words of Van den Hoven (1999 and Van den Hoven and Cushman 1996) we see "a violation of privacy [...] [with] the morally inappropriate transfer of personal data across boundaries of what we intuitively think as separate 'spheres of justice' ". Digital natives[11] have different views of what belongs to the public and what belongs to the private sphere, and these interests conflict with the interests of people that prefer more privacy (International Working Group on Data Protection in Telecommunications 2008). A way to prevent dissemination of information to multiple

[10] www.hyves.nl

[11] Marc Prensky (2005) defines 'Digital Natives' as "today's students [,] [n[ative speakers of technology, fluent in the digital language of computers, video games, and the Internet." He contrasts them with 'Digital Immigrants', the people that adopt and have to learn every new technology.

spheres is using 'blocked exchanges' that separate information from one sphere to another. In one of my weblogs (Riphagen 2008c), I mentioned that the Social Network Site, Genome, claims to implement these blocked exchanges by classifying real-life relationships as family, significant other, business partner, and mapping the information flow based on these relationships.

The separation of information for different spheres touches upon the definitions of 'purpose specification' and 'use limitation' from the Organisation for Economic Co-operation and Development's (OECD) Fair Information Practices (OECD 1980). The international organization that helps tackle the economic, social and government challenges of a globalized economy, proposed eight guidelines for the protection of privacy and transborder flows of personal data (Gellman 2008). These guidelines for the processing of personal information are seen as world-wide standards that private industries should comply with and some countries have even based laws governing their own data processing efforts on it. The purpose specification guideline states that the purpose for which personal data is collected should be specified not later than the time of data collection. This can very well be understood from the point of view of the data subject. She could oppose a copy of her income statement, if it is unclear where this could be used for. This means that the photographer in our example should have notified the subject of his intent to place the photo on a SNS and disclose it to everyone on the SNS who could see it. The use limitation guideline states that the information should not be used for other purposes than the mentioned purpose, unless with consent of the subject or by the authority of the law. If I apply this to the spheres of justice, it means that information collected in one sphere (party) should not be used in another sphere (application for a job), unless by consent of the subject or by authority of the law (law enforcement). This relates closely to the concept of purpose specification, with both concepts aiming at an informed consent[12] by the data subject on all purposes where the identity-relevant information could be used for.

Van den Hoven (2007) also mentions that a fine-grained authorization matrix can serve as a blocked exchange between the different spheres. This fine-grained authorization matrix is translated in computer science to the 'least authority principle'. In an interview with Chris Soghoian (2008b), he explains that this means that when designing an application, each function should get the least amount of information it needs to do its job and not more. This minimizes errors in data systems and abuse of authorities, but will allow applications to perform their functions. The 'least authority principle' is thus a very efficient, safe and easy to implement principle to provide safeguards against malign behavior. With respect to the third-party applications that access the Facebook platform, they get access to much more information than they need (Felt 2008) and therefore information can flow from one sphere (your profile) to another (third-party developers), which could harm you if you do not want these developers to know your sexual orientation, for example.

You are effectively preempted to represent yourself or create your own identity for other people on SNS. Manders-Huits and Van den Hoven (2006) touch upon this in their work. The restriction on one's moral autonomy and the preemption of his ability to create his own moral identity could lead to severe harms, because the data that is relevant to his identity is not under his control. While these may not be accepted by everyone as legitimate reasons to

[12] See the glossary in appendix I for a discussion about informed consent

protect privacy, boyd (2007) finds multiple examples of the problems that can arise when youngsters engage in identity production on SNS. She mentions that in SNS, "bodies are not immediately visible and the skills people need to interpret situations and manage impressions are different". Essentially, she is arguing that users are always restricted in their moral autonomy online because their online presentation is always a digital profile that resembles them more or less. Users do not have as much control over their online moral identity as they have when they meet people face to face. Jenny Sundén (2003) therefore asserts that, "people must learn to write themselves into being." With respect to American SNS, MySpace, boyd mentions that "identity can be seen as a social process that is fluid and contingent on the situation". According to her, teens have always participated in identity formation and the challenges of doing this in public are part of what makes youth grow. However, releasing identity-relevant information in an online mediated environment as part of a social experiment in growing up has different consequences than releasing the same information in a room with friends. The information carries further because it's easily copied, can be stored indefinitely and is searchable. Van den Hoven (2007) adds that teenagers who feel that their privacy is harmed are not so much concerned about the revelation of shameful information, but about their inability to prevent the revelation. In other words "they failed to manage their public face" (Manders-Huits and Van den Hoven 2006). Furthermore, boyd (2006a) mentions that knowing everything about one person changes your perception of them, and eventually your relationship with them. Data-protection laws incorporate the protection of identity management by requiring 'informed consent': the subject of the collection, processing and dissemination of the information should be fully informed by the data processor and fully consent. In an interview, Solove (2008) says that the informed consent that most SNS require does not leave much freedom for the users: "[They are] given a take-it or leave-it choice." These issues are discussed in paragraph 4.4.

The example of the embarrassing photo at the disgraceful party is striking. First, the photo is taken at the party (Solove's information collection), and not many people would judge this as immoral[13]. The picture is posted online (Solove's information processing), and as long as this would happen on a website that allows restricted access in addition to the subject's consent, this also is not very objectionable. The problems start when the photo is disseminated (Solove's information dissemination) to people without the subject's consent. The picture might be disclosed to persons belonging to a social sphere within which the subject does not want the photo to be disclosed (Van den Hoven's informational injustice). Furthermore, by tagging the photo with the subject's real name, the photo can become the basis for scrutiny by future employers (Van den Hoven's information-based harm). Because the users cannot unbundle the services from the SNS, he cannot ex ante prevent the dissemination from happening (Van den Hoven's informational inequality). Finally, if the photo is disclosed to people that the subject does not know, he does not have the freedom to present his moral identity to the people as he wants, as those people already have an image of him (Van den Hoven's restriction on moral autonomy and preemption of moral identity).

[13] Note that, although I say that many people would judge taking a picture at a party as immoral, the Fair Information Practices (Gellman 2008) have a principle 'purpose specification', that states that the purpose for which the data is collected should be specified not later than at the time of collection. The use of the information should be limited to that purpose. In other words, the photographer should have notified that subject of his intent to place the picture on a SNS and the subject should have consented to that, according to FIP.

2.4. How privacy risks materialize: risk analysis

Not every activity that Social Network Sites perform and Solove (2006) identifies as potentially harmful materializes in a harmful incident and resulting damage or harm. Also, it is unclear what the resulting incidents from threats are and what the exact damage is that they inflict upon users. A risk management approach as proposed by Van den Berg (2008) distinguishes a threat from an incident, the resulting damage (harms) and recovery, as depicted in illustration 3. With this approach, privacy risks can be identified and decomposed into the potential negative impact they have (damage) and the probability that they (an incident) may occur. Both of the factors are difficult to identify and therefore I conducted an expert consultation (see chapter 6) and assessed the relative negative impact and relative probability of privacy risks occurring. The experts identified several incidents in which the privacy of users of SNS was compromised. By decomposing the threats that lead to those incidents, options for preventive measures can be formulated, as depicted in illustration 3. This risk management approach is key to identifying and prioritizing privacy harms and coming up with preventive measures to protect users of Social Network Sites. Therefore will the phrases used here be used throughout this thesis.

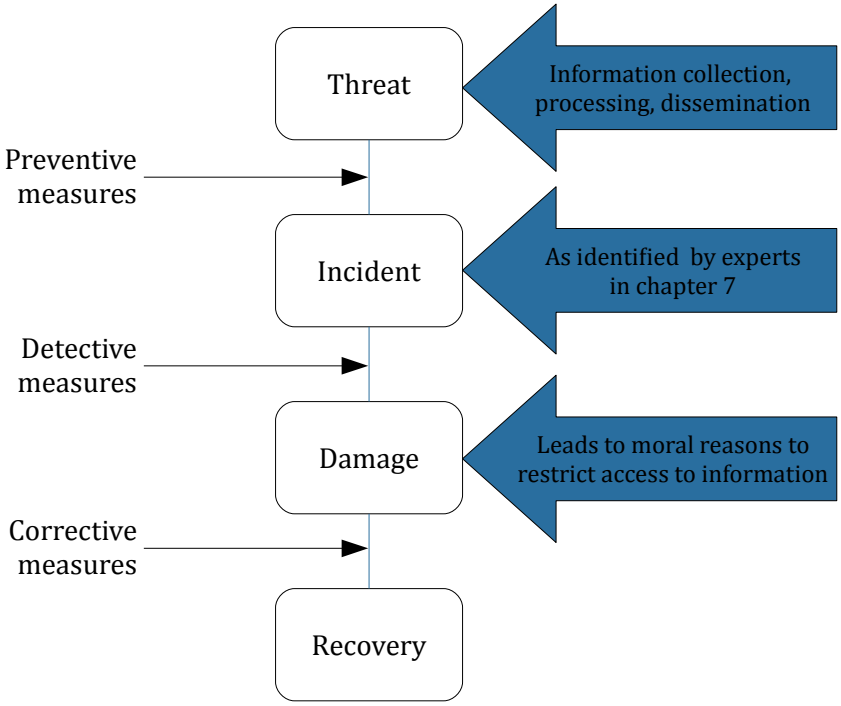


Illustration 3: Risk analysis framework and its relations with this study, as adapted from Van den Berg (2008)

3. Framework: structuring privacy harms in Social Network Sites

Social Network Sites consist of profiles of user-generated information, which are stored in databases and accessed via different functions offered by the website. However, because users of SNS connect with each other and share information via these functions and connections, complexities arise that make it difficult to predict the dissemination of information through the website. And difficulties in the prediction and analysis of privacy risks for users of SNS make it complicated to come up with effective preventive measures.

The complexities in Social Network Sites can be divided in three groups: technical, economic and social complexities. The technical complexities are caused by the complex structures or topologies of SNS. These topologies determine how information flows through the network. For example, an announcement for an event might be placed by a user with relatively few friends in her friends list, but when this announcement reaches a hub (a user with many friends), the dissemination of this announcement through the network will speed up significantly. The economic complexities are caused by the incentives of SNS owners to monetize upon the identity-relevant information that users post online, as they offer the service free-of-charge. Business models and value chains are constructed around these profiles, in an attempt to make money from identity-relevant information. Because of informational inequality users are not always able to get a fair deal. A good example is the collection of identity-relevant information from third-party websites (such as a list of videos you have rented) by Social Network Sites. I refer to the Beacon example in paragraph 3.5 for a more elaborate discussion. Next to technical (topology) and economic (value chain, business models) complexities, the social environment and the social interaction between users of SNS make it difficult to assess what privacy risk can result from uploading identity-relevant information, what their probability of occurrence is and what negative impact they have. Gross et al. (2005) mention that users are “happy to disclose as much information as possible to as many users as possible” and that users perceive SNS as having a great sense of trust and intimacy, while the information can easily be accessed by outsiders. boyd (2004) adds that users present themselves strategically in their profiles. They present themselves prettier, smarter and more interesting to others users, making facts up and posting sexy (manipulated) pictures. These social aspects influence the topology of the network and demand changes in the business models that SNS exploit.

These complexities and trade-offs in different (scientific) fields ask for a more thorough examination in themselves. Furthermore, they are interconnected. For example, the topology of a SNS changes when new partners are inserted into the network as part of the value chain. Those new partners, such as third-party application providers, interact with the users and influence the interaction between users and a SNS. Finally, the government might install regulations, which influences the way SNS conduct business. These complexities can be analyzed on different dimensions. The regulations that govern the activities of SNS are on a different level than the hardware that determines what kind of services can be delivered. During the kick-off meeting of this thesis, we discussed constructing a framework that incorporates all factors that influence privacy in SNS, but describe these factors on different dimensions: from an abstract level of laws and regulations to an as concrete as possible level of why a specific risk to privacy harms a person (Van den Hoven et al. 2008). Such a framework can be used to create a more structured insight into what Social Network Sites are

and how users experience harm when privacy risk materialize. To identify privacy risks in SNS it is important to understand the way these websites work and have insight in the key factors that make up a SNS. This is the first dimension of the framework. The threats for privacy are caused by different activities. Preventing these threats from occurring thus asks for different preventive measures per activity. This is described in the second dimension of the framework. A combination of the activities that threaten the privacy of users, lead to specific privacy incidents, as depicted in illustration 3 in the previous chapter. The outcomes of these incidents, or the exact harms, are described in the fourth dimension. Schematically, the framework looks like this:

1. **Description of the key factors** of Social Network Sites.
2. **Classification of privacy threats**.
3. The various **privacy incidents** themselves.
4. The various **outcomes, damages or harms** of these mechanisms.

The first dimension is the highest level of aggregation and therefore gives the broadest view of what Social Network Sites are. It describes all the various aspects of SNS. This layer details technical, social and economic features of SNS. This layer is essential to identify the incentives SNS have to profit from the information users upload and the regulations and laws they have to comply with. The trade-offs should be made clear here.

The second dimension is based on Solove's (2006) privacy taxonomy. Mueller mentioned during the kick-off meeting of this project (Van den Hoven et al. 2008) that it provides a good framework for breaking down the worries about privacy in SNS into different categories. By distinguishing between different activities that could inflict privacy harms, it is easier to come up with preventive measures. Solove's framework also allows for classification of harmful activities that are considered immoral, but not prohibited by law. The reach of this layer is thus broader than that of a strictly legal perspective.

The third dimension presents the specific incidents that compromise the privacy of users. What activities that make use of identity-relevant information from SNS inflict privacy harms for users? These activities are defined as colloquially as possible. To identify these mechanisms, I have conducted more than 30 surveys and half a dozen interviews with Internet and privacy experts in the USA. Also, I conducted an in-depth case study of American SNS Facebook and draw from earlier research on privacy harms in SNS. For an example of how these activities are described, I point to the example in paragraph 3.5.

In the fourth dimension, I go deeper into the fundamental worries about the various harms resulting from the use of identity-relevant information from users of SNS. This layer deals with the fundamental worries about information on SNS. What could harm people and why does it harm them? Why is there a public outcry when specific features are added to Social Network Sites? This layer mentions the specific reasons why we want to restrict access to the identity-relevant information that we upload to SNS.

These four dimensions allow for a broad description of SNS, which is necessary to identify and describe the broad possibilities of privacy risks occurring through the use of the identity-relevant information available on SNS. Because the framework deconstructs and

describes as ordinary as possible the specific harmful activities and harms, it also allows for policy makers to come up with preventive measure to out a stop to these activities.

3.1. Description of key factors of Social Network Sites

Social Network Sites belong to an emerging set of online communities, which the media call Web 2.0 (Hoegg et al. 2006). These online communities have different business models than older Internet businesses. To describe SNS and their environments in a systematic way, Hoegg et al. (2006) developed the MCM framework. In paragraph 1.4, I explained that the MCM framework was specifically developed to overcome the shortcomings that regular frameworks have in describing new Internet-based businesses and thus is suitable for the analysis of SNS. Furthermore, analyzing Social Network Sites from only one of the various scientific areas determines how the privacy of users is conceptualized. For example, an economic analysis of identity-relevant information as a private good will differ from a moral analysis referring to the 'spheres of access' in a Social Network Site. The MCM framework leaves room for viewpoints and interpretations from other scientific fields, and lets the researcher choose the exact analysis he is using per area of the framework.

The MCM framework is divided into seven generic components, which I have adapted, as mentioned below, to fit more with the specific features of SNS. Adaption of the framework was necessary because Social Network Sites do not charge users for their services, do not deliver physical goods to their users, create value by letting others add applications to their services, and a privately held companies of which not much financial information is known. The components of the framework are:

1. **The societal environment.** This component reflects everything outside the influence of the business model, the external factors. Areas of interest for this component are laws, regulations and the Fair Information Practices and OECD's Guidelines, that should govern the activities of SNS. Also, authoritative agencies such as the International Working Party for Data Protection in Telecommunications and the European Research Agency ENISA have reported on privacy risks for users of SNS and recommended rules to govern these websites.

2. **The features of the medium.** This component deals with the features for transaction and interaction on the medium. As I mentioned before, SNS form specific topologies that influence the speed of information dissemination through the network (Riphagen 2008a). The analysis of the topology of SNS will form the core of this component.

3. **The value chain.**[14] This component describes the value added to the service by the SNS and other actors. The sum of the added value defines the Social Network Site service. Their interrelationships are described by means of an actor analysis as provided by the SEPAM[15] program.

[14] The value chain is the set of activities that a firm employs to create value. For a more elaborate description, see the glossary in appendix I.

[15] Systems Engineering, Policy Analysis and Management program of the faculty of Technology, Policy and Management of Delft University of Technology.

4. The fourth component describes **the (potential) customer** and describes how **value for the customer** is created. This component is added to the third component, the value chain, because the customer is both consumer of other profiles and producer of information and thus part of the value chain.

5. **The specific features of the product.** This component deals with the design of the online community and the way the service is experienced by the users. The GUI (graphical user interface)[16] of a SNS forms an important part user experience. This is also defined by the policies and default settings of a SNS. I argue that the policies (such as terms of use and privacy policy) are part of what could be legally referred to as an 'unconscionable contract'[17]. In this regard, the contract can be assessed in terms of fairness.

6. **The financial flows.** This component describes the earning logic and the various business models that Social Network Sites employ to create revenue. A good example of such a business model is the Beacon example in paragraph 3.5.

7. **The flow of goods and services.** This component describes the processes that create revenue for the company and the processes to deliver the Social Network Site's service to the users. This component is added to the financial flows component because most revenue-generating processes are already incorporated in that component. Furthermore, the component about the specific features of the product describes many of these processes.

The MCM framework will support a structured analysis of Social Network Sites. The framework leaves room for interpretation, but it also has its disadvantages. The MCM framework is initially business oriented, and it is drawn from different business models. It pays less attention to social and human computer interaction factors that determine much of the pay-off structures (see paragraph 4.2) between privacy and social interaction for users. This is incorporated in the component 'specific features of the product', by referring to work done by sociologists as danah boyd. Another disadvantage is that the framework is not validated by other scientists. This is a severe drawback. Because the framework is based on 40 case studies of Web 2.0 communities (Hoegg et al. 2006) and I have adjusted it to fit better with SNS, it is adequate to obtain valid results. Because I incorporate different views in the framework and its usage its limited to structuring, the probability for structural flaws is minimal.

3.2. Classification of threats

Several harmful activities can be performed with the identity-relevant information of SNS users. Solove's (2006) privacy taxonomy provides the basic classification of these activities into information collection, information processing, information dissemination and invasions. As Van den Hoven (2008a) mentioned during the mid-term meeting, the fourth

[16] A Graphical User Interface, commonly abbreviated as GUI, is the graphical representation of computer software that is presented to the user. For a more elaborate description, see the glossary in appendix I.

[17] An unconscionable contract is a specific form of deceptive and unfair contract that has been described as "one which no man in his senses, not under delusion, would make ... and which no fair and honest man would accept" (Leng 2006). See the glossary in appendix I for a more elaborate discussion.

classification in Solove's taxonomy is more of a normative than a descriptive nature. Collection, processing and dissemination are all activities which regard identity-relevant information as the object. Invasions are experienced as such by the data subject, the person whom is the data about, and have him or her as an object. He or she determines whether or not something is an invasion, and it is thus a normative classification. Because I use the taxonomy to classify the privacy harms and not to obtain a moral judgement, I leave the category of 'invasions' out of the classification.

3.3. Privacy incidents

The activities that could harm users of SNS should be described as detailed as possible, so it becomes clear how these activities exactly hurt users. By identifying the exact activity and the harm it does, it is possible to come up with preventive measures. In the study on risks in Social Network Sites conducted by the European research agency ENISA, a similar approach led to the identification of threat scenarios and vulnerabilities. To identify the incidents that are triggered by privacy threats I have asked more than 30 privacy and Internet experts to participate in a survey and interviewed half a dozen privacy experts. I asked them to mention their top three privacy harms for users of Social Network Sites and give priority to the harm that most concerns them. For the methodology of the survey and the data analysis I refer to chapter six.

The privacy harms are classified by the taxonomy proposed in paragraph 3.2. Furthermore, I have asked the respondents to rate the privacy harms on 'probability of occurrence' (how likely is this to happen to a large audience) and 'negative impact' (how severe is the harm derived from this mechanism). Both scales are measured on a Likert-type item, providing quantitive data from a interval scale, more than sufficient for the relative analysis of prioritizing the privacy incidents.

To provide an alternative way to verify the privacy harms that the experts identified, I have asked them to rate to what extent they think the various American tort laws address the privacy issues present in Social Network Sites. I use these outcomes to contrast them with the privacy harms that the experts identified.

3.4. Harmful outcomes or damages

The specific outcomes per harmful activity differ per person. To give a concrete example, the privacy-related incident 'identity theft', acquiring credit by using someone else's credential information -- information dissemination from the perspective of the SNS, has different outcomes for people who have one million euros in their bank account or people who have a thousand euros in their bank account. The outcome might be different, the mechanism and the harm experienced are the same. Refer to the risk management approach in paragraph 2.4 on how privacy threats do not always lead to the same harms for different people.

The outcomes layer of the framework consist of concrete real-world examples of privacy harms for users of SNS, identified by incident, classified by the threat and described in the context of the specific Social Network Site. ENISA (2007) takes a similar approach, describing the outcomes of vulnerabilities on SNS as potential negative impacts.

3.5. Example of the use of the framework

To show the explanatory power of the framework proposed above, I will shortly explain the example of Facebook's Beacon here with the help of the framework.

In November 2007, Facebook introduced its Beacon advertising system, which broadcasts a user's interaction with an advertiser (a third party) to the Facebook friends of that user (EPIC 2008). In terms of the framework, Beacon is described as follows.

1. **Description of key factors**: Facebook, an American SNS, starts a new value chain with its Beacon partners (third parties such as the video rental site, Blockbuster). The goal is to provide viral marketing for the third parties by disseminating the actions of Facebook users on those websites to their Facebook friends. Although there has been some controversy around the sharing of cookies[18] between top-level domains (Van den Hoven 2007), Facebook circumvents using cookies by using HTTP commands and a technique called 'web beacons'[19]. This makes Beacon legal in this respect[20]. The topology of a SNS makes viral marketing very easy and the perseverance of information makes it easy to disseminate information to many friends. There is not much known about the financial flows between the Beacon partners, but it is believed that they participated for free (Solove 2007). They would derive the benefit of viral marketing for free and Facebook would get the benefits from users spending more time on the SNS.

2. **Classification of threats**: Facebook collects information from another website, a third party. This information is processed by Facebook with their proprietary algorithm to determine which of your friends receive your update. Then the information is disseminated to these Facebook friends via your Facebook Mini-feed (see the Facebook case study for a more elaborate description of the Mini-feed).

3. **Privacy incidents:** The harmful mechanisms in Facebook's Beacon and the harms derived from this are plenty. I mention the most obvious ones.

    1. With respect to information collection, Facebook collects the information when a user performs a specific action on the third party's website. Originally, Facebook would provide a short opt-out message on the third party's website (also known as the 'toast pop-up') (EPIC 2008). Facebook's standpoint was clear: if users don't opt-out, we can collect the information. This is a standpoint that is not unusual for marketing organizations that want to acquire identity-relevant information from customers. For an excellent survey, refer to Rotenberg's 'Fair Information Practices and the Architecture of Privacy' (2001). Facebook had an incentive to make this toast pop-up as discrete as possible. If users would switch

[18] Cookies are small text files that websites place on a users computer. These small text files can be retrieved by the webserver later and are therefore referred to as cookies. For examples, see the glossary in appendix I.

[19] Web beacons consist of a small string of code that represents a graphic image request on a web page or e-mail (Krunic et al. 2006). With these small strings of code, information can be collected from one websites and passed on to another website.

[20] The participation of Blockbuster in Beacon is, however, heavily questioned from a legal perspective. Under the American Video Privacy Protection Act, it is illegal to disseminate video rental history of customers to others without their consent (Grimmelman 2007 and Solove 2007).

windows while surfing the third-party website, they could miss the pop-up. The first harmful mechanism: denying informed consent when collecting information. However, after a security advisor performed some research (CA Security Advisor 2007a), which I have confirmed (see my Facebook case-study in appendix A), it became clear that Facebook always collects data from the third party websites, even if you are not a Facebook user. This clearly violates the purpose of informed consent and as users are unaware that the information is collected, it could harm them. You might not want Facebook to know that you have rented movies that do not comply with the mainstream opinion. Because the information that is collected could be used to harm you, this is a case of information-based harm.

2. The information processing happens in a black box with Facebook's proprietary algorithm, that determines to whom the information will be disseminated. Initially, users could not globally opt-out of the feature (EPIC 2008), but after an announcement by their CEO Marc Zuckerberg (Zuckerberg 2007), Facebook provided opt-out of dissemination for all of Beacon or per partner website (Facebook 2008). Legal scholars have discussed (Rotenberg 2001) whether opt-out is the right instrument for data collection as executed by Beacon and if an opt-in feature would not have been more appropriate. From an information processing view, it is relevant that the user has no idea how Facebook determines what information is being sent to which friends: it is a black box. This is at least not in accordance with the Fair Information Practices' use limitation principle, openness principle, individual participation principle and accountability principle (Gellman 2008). Users can be given no security about whether their information is used for other purposes. This is a case of informational inequality.

3. Facebook users' outcries focused on the dissemination of their information to their friends beyond their control (EPIC 2008)[21]. This is a typical case of informational injustice. Users were consuming services or buying products in one sphere (eBay or Blockbuster) and did not expect this information to be disseminated to other social spheres.

4. **Harmful outcomes:** Users of Facebook were upset because they didn't consent to the collection of their information, did not know how the information was processed and could not influence to whom it was disseminated. Imagine buying a book at a bookstore, and if you do not opt-out, the bookstore has the right to tell anyone that you are vaguely connected with what you have bought. One specific harmful outcome worth mentioning was that many users bought christmas presents for their friends online (Beacon was released in November), but all their friends already knew what they were getting. This example shows that the data practices of SNS can have a big impact on our social lives.

The Facebook example illustrates clearly how one new feature of a SNS can be deconstructed into different layers to pinpoint the exact cause of privacy harm. In illustration 4 the decomposition of the Beacon example is graphically depicted. The difficulty with the

[21] Facebook has privacy controls in place. Users can determine if actions from different applications show up in their Mini-feed, noticeable to their friends. I will discuss these setting more in-depth in paragraph 4.3. The Beacon feature initially did not have these settings.

Beacon example is that it encompasses almost all areas of the framework, but the framework does give new insights in how harm erupted from the new feature and why it exactly harmed users.

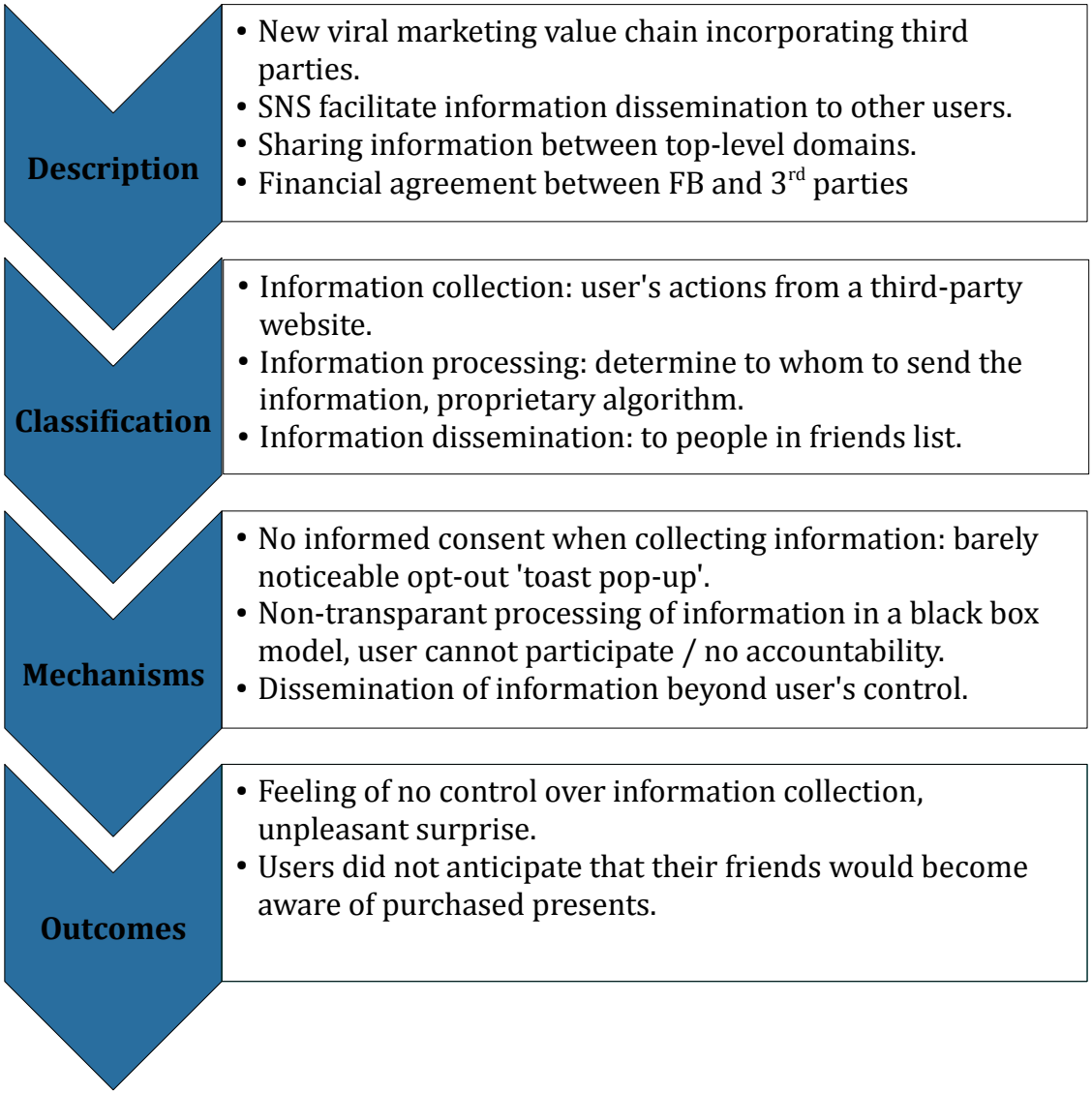


Illustration 4: Beacon example analyzed with framework

4. Description of key factors of Social Network Sites

Social Network Sites are complex web-based communities that bring together many different actors. These actors have different interests. The means that these actors have to achieve their ends are hard to discern because they make us of web technology that is not always visible to end users. The flow of identity-relevant information through Social Network Sites is difficult to monitor, it is therefore hard to restrict access to identity-relevant information, while this is morally desirable. Business models for new web services such as SNS are different from traditional offline business models or the business models based on portals that were common during the dotcom bubble, because they are based on targeted advertising.

Advertisers SNS use the sensitive information of SNS users to create revenue. This information is provided by the users of SNS, as they are both producers and consumers of information. The trade-offs between privacy protection and the creation of revenue should be made clear. SNS are governed by national and international law, as well as what are considered 'best practices' and guidelines for data processing. Hoegg et al. (2006) developed a framework to analyze all these aspects within one framework. As proposed in paragraph 3.1 this framework of Hoegg et al. (2006) will be used to describe SNS from a multi-disciplinary and comprehensive perspective. The framework is depicted in illustration 5. For a detailed description of the framework, refer to paragraph 3.1.

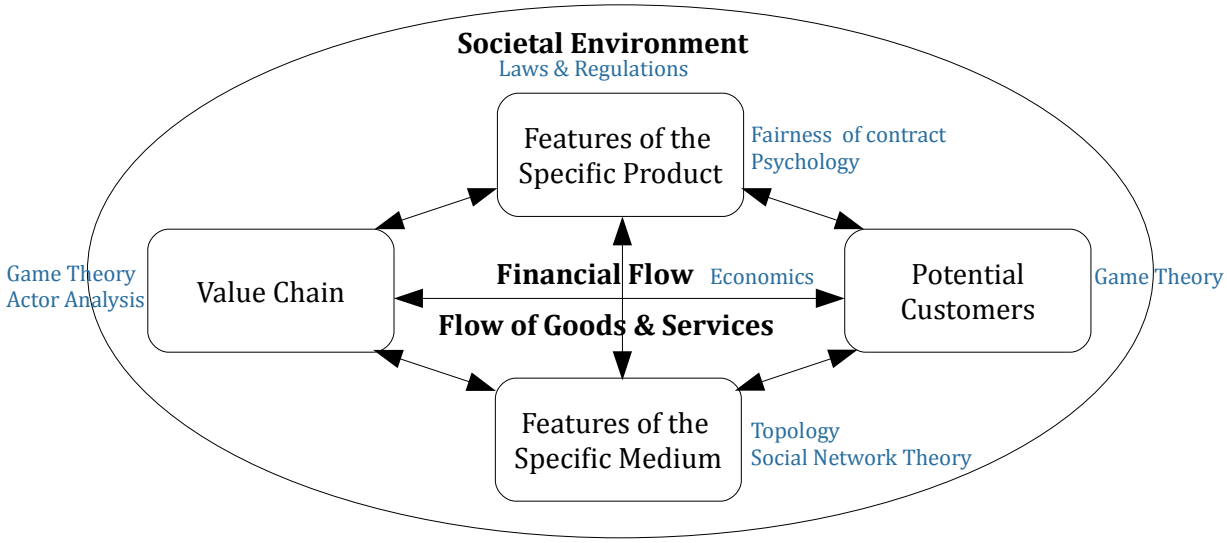


Illustration 5: The MCM Framework used to describe Social Network Sites as adapted from Hoegg et al. (2006), with the different theories used to analyze the specific areas of the framework in blue.

4.1. Societal environment

Web services, like Social Network Sites, have an impact on real life and vice versa (Hoegg et al. 2006). Because SNS have servers in different countries and are world-wide available, they are subject to laws and regulations from different countries. Thus, this paragraph does not provide specific means to enforce compliance with those laws and regulations, but points similarities and differences in these approaches.

American tort laws

This paragraph is too limited to provide an extensive overview of all laws that regulate the behavior of Social Network Sites in the United States. In the survey I conducted amongst Internet and Privacy experts, I asked them, based on advise from Lillie Coney from EPIC, to rate to what extent the different American tort laws address the privacy harms that they identified for users of SNS. The scale I used for the ratings was a 5-point Likert scale, with the values:

1. Does not address.
2. Not really addresses.
3. Addresses a little.
4. Mostly addresses, and
5. Significantly addresses

As the ratings occur on a Likert-type item (Uebersax 2006) I took the mean to summarize the ratings, thereby using steps of 0.25 to identify ratings on multiple levels. For example, a rating of 3,25 would read 'a little' (3) to 'mostly' (4), while a rating of 3,05 would read 'a little' (3).

The experts rated the following American tort laws: the intrusion tort, the publication of private facts tort, the false light tort, the appropriation tort, the right to publicity tort and the breach of confidentiality tort. Although Solove (2006) mentions that "the law of information privacy is significantly more complex [than Prosser's focus on tort law]", I believe that tort laws provide a conceptually sound and proven approach to at least identifying a basic set of privacy harms that these laws should address. Strahilevitz (2004) writes that "tort laws can function as a form of social insurance".

The sample group consisted of 30 experts, the total valid responses to these questions were 20. Because of the low response, I built upon Bots's and Hulshof's model for measuring dissensus in a small sample. I used the techniques and rules mentioned in paragraph 6.5 to measure the consensus of the experts on the rating. In the order of most consensus to least consensus, the experts rated the following tort laws as follows:

- **Breach of confidentiality.** When someone intentionally reveals confidential information about someone else with whom he has a fiduciary relationship, one is liable under the breach of confidentiality tort (Allen 2006). The experts think that this tort law addresses privacy harms in SNS **a little**. The tort fails to apply issues outside the fiduciary or confidential relationship and it is therefore important to determine whether or not there is a fiduciary relationship between a user and the SNS. In paragraph 7.4 I will look further into this issues, but it is evident that the relationship between a lawyer and his client is built on a different type of trust that that between a user and a SNS.

- **Publication of private facts.** Experts agree that this tort law addresses privacy harms in SNS **'a little' to 'mostly.'** This tort imposes liability when (1) one publicly discloses, (2) a matter concerning the private life of another, (3) that would be highly

offensive to a reasonable person and (4) that is not of legitimate concern to the public (Bright 2008). The problem with applying this tort law to Social Network Sites, is that the four conditions are hard to prove with SNS. In a recent case in the Netherlands (Riphagen 2008h) a judge ruled that even a private SNS profile is public with respect to the disclosure of information. This is not a clear matter, and Solove (2006) states that a dichotomous approach of private versus public will not protect privacy adequately. He calls the dichotomous divide 'the secrecy paradigm'[22] and mentions that "[u]nder the secrecy paradigm, privacy is tantamount to complete secrecy, and a privacy violation occurs when concealed data is revealed to others. If the information is not previously hidden, then no privacy interest is implicated by the collection or dissemination of the information". Strahilevitz (2005) has proposed a social network theory of privacy. He claims that this solves the problems concerned with the secrecy paradigm, as it determines which information should be considered private and which not. His approach is more thoroughly discussed in paragraph 4.2. Solove (2007a) writes that his theory could help court in determining when a secret is no longer a secret and the expectation of privacy is diminished. He adds that norms of confidentiality in different networks or social spheres also play a significant role in the dissemination of information. Additionally, what type of information is highly offensive to a reasonable person, especially with users posting sensitive information on their websites[23]? How can judges determine whether or not information is not of legitimate concern to the public[24]?

- **Appropriation.** The appropriation tort creates a liability when someone uses the likeness of another for his own benefits (Allen 2006). For example, a photograph of someone is used for advertising or marketing purposes, without the consent of that person. Experts found that this tort law applies to privacy in social network sites **'a little' to 'mostly'**, but the tort has not been invoked much at court cases about SNS. In paragraph 5.3, I describe how law professors Solove and McGeveran argue that Facebook's Social Ads program is a violation of the appropriation tort, but that the monetary harm is difficult to prove and that the expected financial benefits are low.

- **Right to publicity.** Experts did **not really** think the right to publicity tort addresses privacy harms in SNS. They rated the harm as addressing privacy harms in SNS 'not really' to 'a little'. This tort law states that a person has the exclusive right over the publicity given to his performance (Allen 2006). Although Sundén (2003) mentions that users of SNS 'write themselves into being' and their usage of a SNS can be seen as a performance, this tort law mostly protects celebrities and performers from candid recordings. In paragraph 7.4 I confront this tort law with the privacy harms that the experts have identified.

- **Intrusion.** The experts agreed the least on the applicability of the intrusion tort

[22] The secrecy paradigm relies on the concept of private or secret information to identify privacy harms. If information is available to the public, it is not deemed private and therefore usage of the information does not cause harms in terms of privacy. For a more elaborate discussion, see the glossary in appendix I.

[23] See the part about European law below on sensitive information and the abundance of this type of information on Social Network Sites.

[24] Refer to paragraph 7.4 for a case on the publication of private facts on the Internet.

and the false light tort on privacy harms in SNS. They thought that the intrusion tort does **not really** apply to privacy harms in SNS. This tort creates a liability when someone intentionally intrudes upon the solitude or private affairs of others. Here the problematic distinction between public and private comes into play (Allen 2006). If a users publishes his private affairs on his profile on a Social Network Sites, do others intrude upon him when they view this information? The secrecy paradigm still does not provide a decisive outcome here.

- **False light.** The false light tort addresses giving publicity to a matter that places the other in a false light (Allen 2008). Experts thought this tort law addresses privacy harms in SNS **a little**. If a person disseminates information about the data subject that harms his reputation, without delivering waterproof evidence, he is liable under this tort. Because most dissemination of such information is the data subject self-disclosing information and the false light of the information is hard to proof, this tort law scores low. However, it seems that newspapers can not publish information they find on SNS without a good background check. In a case I present in paragraph 8.2, Amanda Hudson sued newspapers because they based their articles on faulty information from a profile of a Social Network Site.

The American tort laws provide a good way to create a better insight in why privacy threats for users of SNS turn into harmful incidents. For a link to the privacy harms that experts identified refer to paragraph 7.4.

European law

The European Union has comprehensive cross-sectoral data protection legislation, in contrast to the United States. It consists of two Directives on the protection of information privacy (Klosek 2000): the Data Protection Directive (95/46/EC) and the Electronics Communications Privacy Directive (2002/58/EC). The first is a general directive for the protection of data and the second is specifically targeted at ISPs and telecommunication companies.

The **Data Protection Directive** (DPD) protects the rights and freedoms of individuals regarding the processing of personal data and is aimed at harmonizing data protection standards throughout Europe (1995). The most important rules of this directive concern data quality, legitimacy of the processing operations, special categories of data, rights of data subjects and data security (Klosek 2000). The **data quality rule** states that processing of data should be fairly, lawfully, for a specific purpose, and the data must be accurate, up-to-date and not be kept in personally identifiable form for longer than necessary. The CIPPIC complaint, which is elaborated upon in the next paragraph, states that Facebook does not limit the use of the collected information to a specific purpose. The most important conditions for legitimate processing are that "[t]he data subject [..] consented unambiguously" and "[d]ata processing is necessary for the performance of a contract or in order to enter into a contract requested by the data subject" (1995). Social Network Sites argue that users consent to the use of their data by signing up for the service and accepting the privacy policy and the terms of use. The law students from CIPPIC however, mention that this general consent could never be used to replace a specific consent for the processing of sensitive information (Lawson 2008). The interaction between a SNS and a user can be defined as a contract and assessed in terms of fairness, but to enter into this contract the SNS would not need more than an identifier of the

user, such as an email address. There is no necessity to require the submittal of any other information, such as the date of birth, which Facebook does condition the use of its service upon. The European definition of personal data is mentioned in paragraph 2.2, and the EU additionally recognizes **sensitive data**, which is data revealing racial or ethnic origin, political opinion, religious or philosophical beliefs, trade union membership, and data concerning health or sex life (Klosek 2000). This information can be found in abundance on SNS. As Gross and Acquisti (2005) found, more than 50% of the Facebook users revealed dating interests, relationship status and partner and political interests. Under the DPD, Member States must prohibit the processing of this data, except when the subject has given his or her explicit consent. Here we encounter again the contradiction between SNS arguing that they have gotten explicit consent and users that are shocked to find that their preferences are advertised to their friends, as in Facebook's Social Ads. The EU defines a specific category for sensitive data, so the consent requirement is more strict and can not be fulfilled by the general consent a user gives by conditionally accepting the privacy policy and the terms of use. Under the directive, the data subject has the right to know the identity of the data controller, the recipients of the data, and the purpose of the data processing operation (Klosek 2000). According to Klosek (2000), this becomes problematic when websites use cookies[25], small files that websites can place on a user's computer, because these can be used for multiple purposes. For SNS, I argue that this rule is problematic with respect to third party application developers, who are frequently only identifiable by a website address. The Data Protection Directive requires data controllers to **take appropriate measures against the destruction, loss, alteration and disclosure or further (unlawful) processing of data** (Klosek 2000). The CIPPIC complaint mentions the flaws in security with respect to Facebook Mobile and third-party applications (Lawson 2008) and in a weblog, I have examined the loss of more than a hundred Driver's Licenses via a Facebook temporary code glitch (Riphagen 2008g). Indeed, data security is an important concern for SNS, especially with regard to the sensitive information.

Electronics Communications Privacy Directive is an updated version of the Telecoms Privacy Directive (Klosek 2000) and specifically deals with the communication of electronic information, the use of value added services and electronic mail (2002). A value added service is "any service which requires the processing of traffic data or location data other than traffic data beyond what is necessary for the transmission of a communication or the billing thereof" (2002), thus including the types of services that Social Network Sites offer. Although this directive seems directed at telecommunication operators and ISPs mostly, as its predecessor (Klosek 2000), the sections on directories of subscribers and unsolicited communications might apply to Social Network Sites. If we see the profile and network of a SNS user as a directory of subscribers, SNS operators should obtain additional consent for activities other than searching for contact details of the subscriber. The directive mentions that "Member States may require that for any purpose of a public directory other than the search of contact details of persons on the basis of their name and, where necessary, a minimum of other identifiers, additional consent be asked of the subscribers." (2002). The directive also mentions that a data collector has to get prior consent for unsolicited communication. Furthermore, a company that has customers should provide opt-out options for every email it sends to its customers. In the case of Facebook Beacon, this opt-out options

[25] For a more elaborate explanation of cookies, see the glossary in appendix I.

is now available per instance.

European law addresses the privacy risks of SNS from two directives, however no activities to enforce compliance of SNS with the directives is known at this time.

Canadian law

In May 2008, the Canadian Internet Policy and Public Interest Clinic filed a complaint with the Canadian Privacy Commissioner against Facebook for violating Canadian privacy law on 22 occasions (Shiels 2008). The Ottowa-based privacy research group is affiliated with the local University and students examined Facebook's practices to file a 36-pages complaint, which focuses on the "unnecessary and non-consensual collection and use of personal information by Facebook" (Lawson 2008). Canada is known for having one of the broadest privacy protection laws, the Personal Information Protection and Electronic Documents Act (PIPEDA). CIPPIC found that Facebook did not comply with this Act in the following occasions (Lawson 2008):

- Facebook **does not identify all purposes** for which it collects users' personal information. For example, when users sign up for Facebook, they have to provide their date of birth. Facebook claims it needs the date of birth to verify that someone is older than 13 (the minimum age for use of the site) and to facilitate its birthday application. However, it is a severe action to ask for the date of birth as condition to join the website. Also, a user might not want to use the birthday application, and this unbundling of services is made impossible by the conditional access.

- Facebook **fails to notify and obtain informed consent** from users for the uses and disclosures of their personal information. The default settings of Facebook share all personal information of users with their friends and all personal information except contact information with 'all of my networks'. Users can only become aware of this sharing of their information when they go to their privacy settings. By default, users are not aware of this use and disclosure and can therefore not knowingly provide their informed consent.

- Facebook **fails to obtain express consent to share user's sensitive information**, especially with respect to the photo album, that is by default shared with everyone, even if the user's profile is only visible for friends. User's non-friends can easily view photographs and the associated comments, without the informed consent of the profile owner. The identification of user in a photograph is sensitive information, as it can be used by stalkers and for re-identification on other websites. Acquisti and Gross (2005) were able to re-identify profiles from SNS Friendster with webpages from their University by using a commercial face recognizers. This is an easy solution, which diminishes the probability of having an anonymous profile on dating sites or other Social Network Sites. The default sharing of photos with everyone on Facebook could therefore lead to severe privacy harms. Other types of sensitive information that Facebook shares are age, hometown and school. Users may restrict access to some of this sensitive information by using the privacy controls, but it will still be used for Social Ads.

- Facebook **fails to meet the requirements for valid opt-out consent.** According

to the Privacy Commissioner of Canada, opt-out consent is only valid when users are informed about what happens with the data when they do not opt-out. As mentioned above, the users are not informed about what happens with their information if they do not opt-out.

- **Users can not easily withdraw consent** to the sharing of information they uploaded earlier. Facebook users can not opt-out of Facebook's Social Ads, so the information in their profile will still be used for Social Ads, even if they have made their profile visible to only friends.

- Facebook collects more information than that is necessary for its stated purposes, in other words, it **does not comply with the 'least authority' principle**. CIPPIC mentions that Facebook states its purpose as 'Social Networking' and collecting information to provide targeted advertisement is not necessary to provide social networking services to the users (Lawson 2008). Also, Facebook mentions in the privacy policy that it reserves the right to collect information from other sources than the website (see appendix A - Facebook case study) and because it does not specify when, how and from which sources it collects this information, it is unclear for which purposes it will use the information. As the user is not informed, he can not be expected to give any consent to this collection of data from other sources.

- Facebook is **not upfront about its advertisers' use of personal information** and the control that users have over their privacy settings. By default, users are included into Facebook's Social Ads[26] and Beacon. Users can opt-out of Beacon, but not opt-out of Social Ads. CIPPIC states that Facebook conditions the access to its network on the dissemination of profile information to facilitate Social Ads and therefore violates PIPEDA (Lawson 2008). This conditioning of usage of the service by such invasive dissemination of personal information is unfair with respect to the user, who might want to opt-out of this, and still be able to use the service.

- Facebook **conditions the use of its services** with the collection, processing and dissemination of users' information. Not only can users not opt-out of Social Ads, users can only add third-party applications if they consent to sharing all their information and the information that they can see of their friends with the third party application developer. Facebook also does not advise users that withdrawing the consent to access all their information will lead to the deletion of all third-party applications on the user's profile.

- When a user leaves Facebook, **the information from his profile is not destroyed, but retained** by Facebook. A Facebook user can deactivate his account by clicking on a button under the account settings, but deletion of the profile is only possible when the user sends an email to Facebook. Facebook does not inform the user about what happens with the information that Facebook retains when the user deactivates his account. Principle 4.3.8 of PIPEDA states that an individual may withdraw his or her consent at any time to retention of personal information and

[26] Facebook Social Ads shows advertisement targeted at your profile (Lawson 2008). Beacon is a variant on the Facebook Social Ads program that shows advertisements based on actions users perform on external websites of Beacon partners.

principle 4.5.3 states that personal information that no longer is required to fulfill the identified purposes should be destroyed, erased or made anonymous (Lawson 2008). As the user has opted out of social networking, Facebook can not retain the information for that purpose anymore. Also, Facebook should provide minimum and maximum periods for the retention of the data under principle 4.5.2 of PIPEDA (Lawson 2008). Facebook keeps profiles of deceased persons active under a special memorial status. This is a new purpose and thus requires the consent of the user.

- Facebook **does not protect the profiles of its users adequately** to prevent unauthorized access to the information. Facebook does not monitor the quality or legitimacy of third-party applications, which makes it easy for hackers to abuse the Facebook Platform API to get access to users' profiles and even change settings of the profiles of these users. Also, when users login to the mobile version of Facebook, Facebook provides them with a cookie with login credentials that has no expiration date. Any device used to access the mobile platform of Facebook can therefore indefinitely be used to access the user's account, creating huge security risks.

- Facebook **does not explain their policies and practices on the sharing of users' information** with third-party advertisers and application developers. When users add an application of a third party to their profile, they click on a check box which allows the third-party to access all the information about the user and the information that the user can see of his friends. Christopher Soghoian (2008a) and Adrienne Felt (2008b) found out that the third-party application developers get access to much more information than they need to make the application work and that these applications get access to profiles of the friends of users, without those friends' consent. The extent to which personal information was shared only came to light because of the research of Felt and Soghoian. This does not conform with PIPEDA or the earlier mentioned 'least authority' principle. Furthermore, Facebook has no way to govern the secondary use of this information by the third parties (Soghoian 2008b) and third parties could store the information indefinitely or sell it.

- Facebook **adds new features** to the website, which use the information that users already posted on their profiles, **without adequately informing users and asking for their consent**. PIPEDA (principle 4.2.4) states that if a data collector uses information for a new purpose, this purpose should be identified prior to use and user consent is required (Lawson 2008). By reserving the right to add new applications and modifying the terms of use, Facebook violates this article and the OECD's 'purpose specification' guideline[27].

- Facebook **fails to obtain consent for collecting information of non-users**. Users of Facebook can tag photos of non-users with their name, without the consent of these users. Furthermore, when signing up for Facebook, users are asked to provide Facebook with access to the contact lists of their email software (Facebook Friend Finder, see appendix A - Facebook Case Study), without the consent of these users. Those users get invitations and upon receiving newer invitations, all the old invitations are listed, clearly showing the Facebook retains information on non-users (Lawson

[27] See next paragraph on the OECD's Fair Information Practices guidelines.

2008). In the case of Facebook's Beacon, information about actions of non-users on Beacon partners' websites is always sent to Facebook (refer to paragraph 3.5).

- Facebook uses technology to detect anomalous behavior of users, but **fails to notify users of this.** This is in violation with principle 4.8 of PIPEDA, which obliges Facebook to make its policies related to the management of information readily available (Lawson 2008).

- Facebook **misleads users with deceptive practices.** It represents itself as a Social Network Site and disguises clear marketing activities, like Beacon, as social activities. Beacon is an advertising mechanism paid for by third party advertisers. But this stream of revenue stays very covert from users (Lawson 2008). Also, one of Facebook's two core principles is that "you should have control over your personal information" (Facebook 2007a), but the level of control the users have over the information that Facebook shares with third parties is minimal. As CIPPIC mentions: "Facebook needs to stop promoting itself as a model of the real world and needs to be upfront about the real limitations on the level of control available to users." (Lawson 2008)[28].

The CIPPIC complaint mentions many violations of Canadian privacy law by Facebook. The Canadian Privacy Officer, Jennifer Stoddart, will investigate the complaint and start mediating initially to enforce compliance with PIPEDA. If all fails, the Officer can seek court injunction to resolve the issues.

OECD's Fair Information Practices

In 1980, the OECD erected eight guidelines for the handling of personal data. Developed by a group of government experts, the Paris-based organization devised the 'OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data' (Gellman 2008). Marc Rotenberg (2001), an influential scholar and activist on informational privacy issues, argued that privacy protection in information law is "generally understood as the enforcement of Fair Information Practices". The guidelines are as follows (Gellman 2008):

- **Collection Limitation Principle.** There should be limits to the data collection, data should be collected by **lawful and fair means**, and with **knowledge or consent** of the data subject. Facebook clearly violates this by obtaining information from third party websites through the Beacon program, even if users are not members of Facebook, opted out of the feature or are not logged in to Facebook (see appendix A - Facebook Case Study).

- **Data Quality Principle.** Collected data should be **relevant for the purpose** for which they are to be used **accurate, complete and kept up-to-date**. As mentioned in the Facebook case study, third party application developers get in 91% access to more information than they need to make their services work. It is clear that not all data is relevant for the purpose of making the application work. Banks and insurance

[28] In an interview, John Verdi characterized this dishonest behavior as follows: "[t]hey cultivate the view that you can submit information to a SNS and control where the information goes. [..] In the reality this information is available to many individuals, many organizations and many entities in a variety of ways in a very uncontrolled, unmonitored and unaudited way." (Verdi 2008)

companies that collect information are regulated and restricted on the collection of information that is not necessary for their primary business functions, see paragraph 9.3.

- **Purpose Specification Principle.** The **purpose** for which the data are collected **should be specified** no later tan at the time of data collection and the subsequent use limited to the fulfillment of that purpose. As the CIPPIC complaint mentions, "Facebook does not adequately explain why users must provide their date of birth in order to use Facebook." (Lawson 2008). To check whether someone is older than 13 years, the minimum age for using Facebook, other means can be used. Because Facebook does not adequately provide the purpose of collecting the date of birth, it violates the purpose specification principle.

- **Use Limitation Principle.** Data should **not be used for other purposes** than in accordance with the **purpose specification principle**, except with consent of the data subject or by authority of law. A good example of where Facebook does not comply with the use limitation principle is the Social Ads program, where profile pictures and information of users is shown to other users combined with the products they purchased. This example is discussed more elaborately in paragraphs 5.3 and 8.3. It is important to note that users did not upload their profile picture and their first and last name to be used as an endorsement for a product. They uploaded their picture, first and last name, so (potential) friends could recognize them. Although users can opt-out of Social Ads, the user never gave his consent to the use of the information for this purpose.

- **Security Safeguards Principle.** Personal data should be **protected by reasonable security safeguards.** As mentioned in the Facebook case study, users can access Facebook via their mobile phone. The CIPPIC complaint mentions that Facebook sets a cookie on the mobile device that seems to have no expiration date (Lawson 2008). Since May 30, 2008, Facebook seems to have changed the expiration date on the cookie, as it was now set for three months (see Facebook case study). However, this period of time is much longer than for non-mobile access, where the login cookie is set until the session ends. Another cookie reminds your email address for non-mobile access for three months and 25 days. Reasonable security safeguards mean that Facebook should have opted for the same kind of safety and security settings for its mobile service as for its non-mobile service.

- **Openness Principle.** There should be **a general policy of openness** about developments, practices and policies with respect to personal data. However, as the CIPPIC complaint mentions, Facebook does "reserve the right at [their] sole discretion, to change, modify, add, or delete portions of these Terms of Use at any time without further notice." (Facebook 2007a). Especially with respect to the introduction of new services, Facebook expects the user to read the changes in the terms of use or privacy policy. A cynical view on Facebook's openness entails that updates of your profile are sent to all your friends and third party application developers, while you have to manually check for any changes in the privacy policy or the terms of use.

- **Individual Participation Principle.** An individual should have the right to a) **obtain information about him from a data controller**, b) within a reasonable time

and in a reasonable manner, c) to be given reasons if such a request is denied and d) be able to challenge data relating to him. Up until now, I have not encountered any function that enables users to download all the information that a SNS collects about them. After my appearance on a Dutch national radioshow (Radio Online 2008) in which I stated that Hyves should be more transparent about the information they collect about users, they issued a statement saying that all the information that Hyves collects is on the profile of the user. However, it is easy to prove that this is a false statement. Just by looking at the cookies collected by my browser, I could see that Hyves stores information about the initial status of the Hyves chat, the website that you visited before you went to Hyves (referral) and two other cookies that facilitate Google's Urchin Traffic Monitor (UTM). The latter tracks unique website visitors (Google 2007) and Hyves obviously has incorporated this service to gain information about unique visits, information about you they do not store in your profile.

- **Accountability Principle.** A data controller should be **accountable for complying with measures**, which give effect to the principles stated above. As mentioned in the Facebook case study, a third-party application developer from India created Scrabulous, a game very alike Scrabble, that infringes on the intellectual property rights of the owners Hasbro and Mattel. Although the Developers Terms of Use (Facebook 2007c) from Facebook mentions that any use of the Facebook Platform that infringes upon intellectual property rights is illegal, Facebook undertook no action to resolve these issues.

Other regulations

Several organizations that promote the protection of (online) privacy investigated Social Network Sites, identified privacy risks and made recommendations to minimize these threats. These recommendations influence the societal environment, and in particular the actors that have regulatory power over SNS.

The International Working Group on Data Protection in Telecommunications (IWGDPT) is a group chaired by the Berlin Commissioner of Data Protection and consists of representatives from worldwide Data Protection Authorities and other bodies of national public administrations, international organizations and scientists (Berliner Beauftragter für Datenschutz und Informationsfreiheit n.d.). At their 43rd meeting in Italy, the IWGDPT drafted the 'Rome Memorandum' on Privacy in Social Network Services (International Working Group on Data Protection in Telecommunications 2008). In this report, the IWGDPT identifies privacy risks for users of SNS, which I will treat in paragraph 7.6, and guidelines for regulators, owners and users of SNS. The 23 guidelines can be categorized in five groups: **providing tools** that help users protect their identity-relevant information, a **better regulatory framework** for SNS and complaint handling, **more transparency** from SNS on what data is required and collected for processes and about data breaches, **more priority for security** on SNS and user empowerment by **educating users** about their rights, the activities of SNS and making users aware that they should take more responsibility (International Working Group on Data Protection in Telecommunications 2008).

The European Network and Information Security Agency (ENISA) published a position paper on the security issues of Social Network Sites in October 2007 (ENISA 2007). In this

paper, ENISA identifies the most important privacy threats to users and providers of SNSs and offers policy and technical recommendations to address them. ENISA identified threats in four groups: privacy-related threats, SNS variants of traditional network and information security threats, identity-related threats and social threats. The European agency came up with recommendations for government policy, provider and corporate policy, technical issues and research and standardization.

The **privacy-related threats** that ENISA identifies are digital dossier aggregation, secondary data collection, face recognition, Content-Based Image Retrieval[29], link-ability from image metadata and difficulties with account deletion. It considers SNS spam, cross site scripting (XSS)[30] and SNS aggregators[31] as **variants of traditional network and information security threats**. The **identity-related threats** that ENISA sees are spear phishing,[32] using SNSs and SNS-specific phishing, infiltration of networks leading to information leakage and profile-squatting and reputation slander through ID theft. Furthermore, ENISA recognized the following **social threats** of stalking, bullying and corporate espionage. Where appropriate, I will mention the risks that ENISA identified in this thesis.

To combat these privacy threats in Social Network Sites, ENISA recommended the following **government policy recommendations**: encourage awareness-rising and educational campaigns, review and interpret the regulatory framework, increase transparency of data handling practices and discourage the banning of SNS in schools. With respect to **Social Network Site operators**, ENISA recommends to promote stronger authentication and access control where appropriate, implement countermeasures against corporate espionage, maximize possibilities for abuse reporting and detection, set appropriate defaults and offer convenient means to delete data from a user profile completely. The **technical recommendations** that ENISA makes are to encourage the use of reputation techniques, build in automated filters, require consent from data subjects to include profile tags in images, restrict spidering and bulk downloads, pay attention to search results, address SNS spam and address SNS phishing. Finally, ENISA also **recommend actions on research and standardization**, to be specific to promote and research image-anonymization techniques and best practices, to promote portable Social Network Sites and to do research into the emerging trends in Social Network Sites. In the rest of this thesis, I will touch upon the recommendations that ENISA makes on several places.

4.2. Features of the medium relevant to SNS: Topology of SNS

Social Network Sites consist of profiles of users that are connected with other profiles of

---

[29] Content-Based Image Retrieval is an emerging technology that can match features, such as an identifying aspect of a room (a painting) to a database with location information, increasing the probability for identifying a user's location (2007).

[30] XSS consists of including scripts that are running on one website to manipulate data on other websites. The scripts on the first website are not governed or monitored by the second website. For example, on many SNS users can post information on their profile in HTML

[31] Social Network Site aggregators are websites that collect, process and disseminate information from multiple SNS platforms on just one website.

[32] Phishing attacks involve the creation of a fake website or email, which resembles a real-world website or email (social engineering), to abuse the users trust in this website and inquire for logins and passwords.

users. These structures form networks that can be modeled and analyzed with graph theory. The profiles are the vertices (nodes) and the links between the profiles are the edges (links). The topology of such a network, in other words how many vertices and how many edges a network has and how those are connected, influences the way information disseminates through the network (Riphagen 2008a). Strahilevitz (2004) notes that we all reveal facts to our friends in social networks, and consider those to stay secret. However, he notes that the structure of the network, as well as the culture of the network and the content of the information influence the speed with which information disseminates through a social network. Also, Social Network Sites use proprietary algorithms to determine to whom to disseminate specific information and create value from the network of their users. The structure or topology of a SNS is thus very important to determine potential privacy risks.

The path between any two users can be measured, and the set of all these paths has a minimum and a maximum. The average shortest path is the mean of this set. I found SNS to have an average shortest path of 5,38 (Riphagen 2008a), which means that any user can be reached within 6 steps. The longest shortest path, the upper bound on how many steps it takes to reach a certain person, is 7 steps or more. This is also called the diameter of the network and a measurement for how fast information travels from one part to another part of the network. The maximum diameter that I found in SNS is 15,33 steps (Riphagen 2008a) It has to be noted that SNS vary in their practices of establishing paths between users. As Hyves only allows the information to flow between users when you have a bilateral agreed upon connection or have a public profile, Facebook allows anyone in your network (for example the Washington DC network) to access your profile (Lawson 2008). This results in information flowing much easier through the Facebook network.

The clustering coefficient[33] determines how well friends of you are connected to each other. In other words, if you know friend A and friend B, what is the probability of those two friends knowing each other? Strahilevitz (2004) mentions that social networks are highly clustered. A high clustering coefficient (meaning high clustering) and a low average shortest path are signals that the network has small-world effects (Strogatz 2001). Small-world effects indicate that any user of the social network can connect with anyone else in a few steps. Stanley Milgram found out that anyone in the world can connect with any other person within 6 steps (Strahilevitz 2004). Note that the average shortest path in Social Network Site is lower, namely 5,38. Two indicators of these small-worlds effects are the average degree and the assortativity of the network (Riphagen 2008a). Fu et al. (2007) find that SNS users have on average 35,8 connections or a degree of 35,8. However, Strahilevitz (2004) mentions that Social Network Sites incorporate users that have an enormous amount of connections, the super nodes. Once information reaches a super node, the dissemination speeds up significantly. The distribution of the degrees of users of SNS has a power-law tail (Smith 2002), proving that a few nodes have very many connections and many nodes only have a few connections. The assortativity[34] of a network measures whether these super nodes are more likely to connect with each other than with other users. This would increase the speed of information dissemination significantly. This is called preferential connection and it leads to growth of the network and high clustering. The power-law tail of SNS seems to point at

[33] For the mathematical definition of the clustering coefficient, refer to the glossary in appendix I.

[34] See appendix I for a more elaborate explanation

preferential attachment, however this has not been proven.

The connectivity of users and the structure determine the speed of information dissemination through the SNS. Solove (2006) mentioned information dissemination as one of the activities threatening privacy. Strahilevitz (2004) proposes to look amongst these lines to define expectations of privacy. He mentions that people disseminate information to a small group of friends, but expect the information to stay private and they expect that this information does not disseminate outside this group of friends. These expectations are not covered by the traditional 'secrecy paradigm' (Solove 2006), because this would render the information already public, because it is not private anymore. This is not in accordance with the preferences of the users. Strahilevitz turns to Social Network Theory to explain why data subjects can reasonably expect information that they disseminated to a group of friends to stay private.

The structure and culture of a network and the content of the information determine how fast information travels through a network and what expectations of privacy users can expect to have. Strahilevitz (2004) writes that "there is always the potential that information that any persons discloses to any other person will spread to the entire world", but that structural holes[35] prohibit this dissemination. The structure of a SNS, with its super nodes, closes these structural holes, creating a high level transitivity. These super nodes function as weak ties that gather information from one social sphere and disseminate it to another. In the case of SNS, people that you do not really know but have added to your friends list, can fulfill this function of weak tie and super node. However, not all information is easily communicated via weak ties. Complex knowledge, such as a combination of knowledge from different fields, is difficult to convey because it requires are more intimate relationship (Strahilevitz 2004). However, the profiles on SNS are quite appropriate at conveying complex information. By sending just one URL that links to a profile, users can easily disseminate complex knowledge. The technology that SNS offer facilitates information dissemination greatly.

Cultural factors that influence the dissemination of information through a social network are moral constraints, the type of information that is disseminated and the ability to determine what information other network members deem relevant. SNS encourage information dissemination and it is easy to find what information other members find interesting. For example, by clicking on a specific book in Facebook that is of your interest, you will see all the other members of your network that also listed that book. This could be a great advantage for users of SNS, as it is easy to convey information to the right people, but equally it leads to privacy risks. The more interesting a specific piece of private information is, the less likely it is to degrade as it passes through the network (Strahilevitz 2004).

Strahilevitz's analysis is interesting, because in Social Network Sites you are defined by your social network or who you know. Even if you put your profile on only visible for friends, you can be identified and profiled by the connections that your friends have. The people in your network also have preferences. Aggregation of their preferences can be used to identify you. This has implications for your moral identity, as SNS are still able to profile you based on your friends' preferences and your connections with them. For example, you might share your taste in music and books with your friends, and even without disseminating your personal

[35] Structural holes are defined by the absence of connections between two nodes. Strahilevitz (2004) describes them as a lack of effective ties, and therefore information can not flow from one person to another..

preferences to others, a digital dossier of your network can reveal those preferences. By accumulating information from incomplete profiles and placing this in the context of the social network, SNS are able to profile users. As mentioned before, it is known that Social Network Sites use algorithms to decide to whom which information is being disseminated. As these algorithms are proprietary, their workings are unknown. However, inference with information that is not uploaded by the user or with future information could pose privacy risks, because this inference is outside of the control of the user. Users are also defined by the social network they are a part of, and many users identify themselves with groups or networks by becoming members. Facebook has groups based on high school or university, but also groups for fans of specific bands. Even though a user might not be a member of such a group, if many of his friends are, the chance is high that he will be related to that group as well. Although it is unknown on which scale this happens, the police in Amsterdam has issued a press release stating that it will use SNS to profile suspects and their networks. This inference with social network information can harm users, especially when the information is incorrect. The feeling of having no control over one's own information especially appeals to our feelings of morality.

4.3. Value chain and customers

The value chain[36] of Social Network Sites is very different from the linear value chain that Porter (1985) described. Many actors contribute to the value chain: users upload identity-relevant information of themselves and others, advertisers deliver advertisement on user profiles, and ad networks as Doubleclick provide specific users with targeted ads. Furthermore, third party application developers develop and provide applications and partners provide the Social Network Sites with specific information about their users' behavior in exchange for free viral marketing (also known as Facebook's Beacon program).

The actors involved in the value chain of SNS all have different interests. Their incentives for economic benefits do not always contribute to an adequate restriction on the identity-relevant information that users post on Social Network Sites. Users of Social Network Sites have good moral reasons (see paragraph 2.3) to restrict access to the information they upload, but they do not have the right instruments to restrict this access. I will treat the uploading of identity-relevant information by users as given, and focus on the other options that there are to restrict access to this information and protect privacy in SNS. The actors that are involved in the value chain of SNS are the Social Network Site owner, the users, third party application providers, advertisers, advertising networks, partner websites, government and regulators, and activists and researchers. Each of these actors has their own influence on and place in the value chain. The activities from the privacy taxonomy of Solove (2006) provide a good way to depict the relations of the actors graphically. Part of the actors are involved in information collection, by providing the SNS with information, uploading identity-relevant information about themselves or providing content in the form of advertisement or applications. The Social Network Site processes all the collected information, with the use of proprietary algorithms (see paragraph 2.2) that determine which information is disseminated to whom at what time. Other actors receive information disseminated from SNS or request information from SNS to provide services to the users. Examples of the latter are third-party application

[36] The value chain is the set of activities that a firm employs to create value. Porter (1985) identified the following areas that make up the value chain: inbound logistics, operations, outbound logistics, marketing & sales and service.

providers and advertising networks, who both provide users with services based on the information in their profile. The government and privacy regulators govern the value chain of SNS and regulate the activities that SNS employ. Activists and researchers turn to SNS for a wealth of information or highlight the privacy risks that derive from using these services. These relations, with the flows of information illustrated by orange arrows, are depicted in the illustration below.

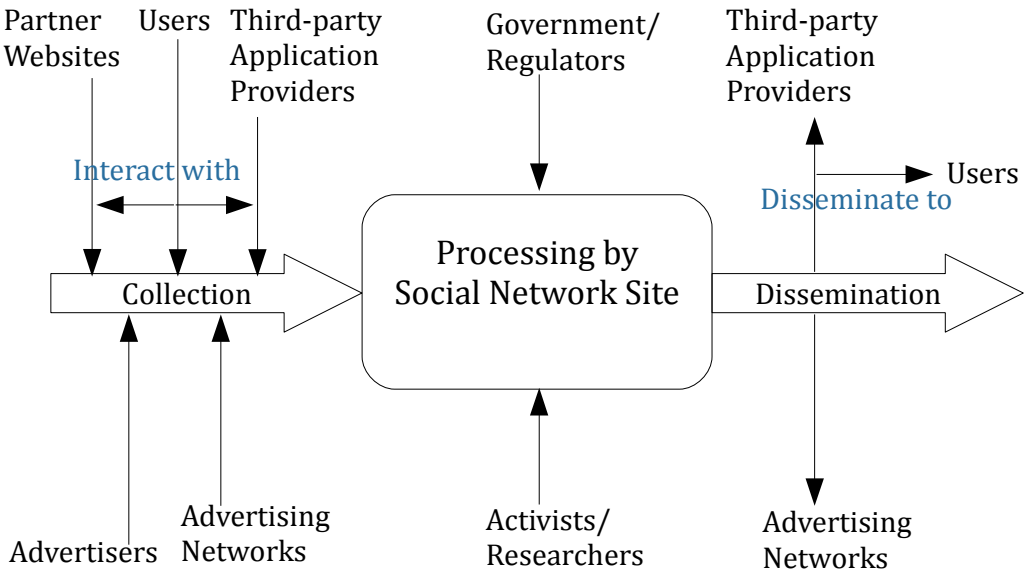


Illustration 6: the actor network of Social Network Sites, depicting the various actors that are part of the value chain.

This illustration of the actor network and value chain of SNS is a momentarily image, that changes easily, for example when SNS introduce new services. With the addition of Social Ads and Beacon, Facebook incorporated many more actors in the value chain. Note the flow of information between the users and the partner websites, who provide information to the SNS about the behavior of users on their websites. Users also provide third-party application providers with information, namely about which applications they want to add to their profile.

In appendix Q, I provide a list of all the actors that are in the value chain of SNS or have the ability to influence this value chain. Social Network Sites and private organizations involved in the value chain want to maximize profit by increasing the amount of users that use their applications, increasing the exposure of their advertisements by viral marketing on a SNS or the optimal targeting of advertisements to users. However, users want to derive benefits from uploading information to a SNS and interacting with their friends, without the negative consequences of privacy harms. The government is concerned with finding a good balance between the value creation by SNS (and thus economic revenue and growth) and the protection of the civil and privacy rights of their citizens. Researchers and activists are mostly concerned with the complexity of privacy issues in SNS, the lack of attention from a policy perspective and the gaps in regulation.

The perceptions of the problem vary greatly between SNS and their users. While Social Network Sites put the emphasis on better user education and a greater array of privacy controls, thereby effectively putting the blame on the users, users claim to not have enough

control over the identity-relevant information they upload to SNS. As mentioned in the CIPPIC complaint (see paragraph 4.1), users do not have control over all the information collection, processing and dissemination and information is collected by SNS without the consent of users. By giving users more privacy controls, Social Network Sites just partly solve the problem, as the controls do not address all information collection, processing and dissemination and it is mostly unclear what the exact outcomes are of using a specific control. As Felt (2008a) mentioned in an interview:

"When you get to really fine-grained controls, it gets confusing to users. And I think [the controls] that [..] Facebook has [are] already really confusing. [Facebook is] trying to have fine-grained controls, but I need to show to a lot of people how to set up their privacy controls because they don't understand what's going on. [..] The more fine-grained your controls are, the more settings there are. A full page of checkboxes would be very confusing to users."

Furthermore, the interests of users and Social Network Sites are conflicting. Users want to have control over their information, while benefitting from the social contacts that SNS facilitate. SNS want to maximize their profit and therefore promote the free flow of identity-relevant information, because this will increase the demand for and thus the monetary value of the information. Users do not have any means to limit the negative consequences (privacy harms) from these incentives, except for calling for the restriction of these activities and influencing the public opinion. The government and regulators could develop new legislation to find solutions which does right to both interests.

The incentives of Social Network Sites and users can be illustrated by showing the options they have and the pay-offs of these options in a matrix. An easy definition of the utility that both actors derive from Social Network Sites gives insight in how users end up in a disadvantage. The utility functions are as follows:

User utility from SNS = Benefits from increased social contact - costs from privacy incidents.
SNS owner utility = Monetary benefits from amount of identity-relevant information on SNS - costs from operations and complying with regulations.

For the sake of simplicity and because not much is known about SNS incentives, I have reduced the options that users and SNS have to two for each. Users have the option to (1) not join a SNS and (2) join a SNS. Social Network Sites have the option to (1) not restrict access to identity-relevant information and for example share it with advertisers for profiling and (2) implement measures that restrict access to the information that users upload. In table 1 their pay-offs are qualitatively depicted between brackets on a scale from -2 to +2. The pay-offs refer to the utility functions mentioned above.

	SNS restricts access to information	**SNS leaves access to information open**
User does not join SNS	User: (0) - (0) = (0) SNS: (0) - (1) = (-1)	User: (0) - (1) = (-1) SNS: (0) - (0) = (0)
User does join SNS	User: (1) - (0) = (1) SNS: (1) - (1) = (0)	User: (1) - (2) = (-1) SNS: (1) - (0) = (1)

Table 1: options with pay-offs for Social Network Sites and users.

The model is based on game theory, a science that investigates options and outcomes of multi-actor situations in the terms of alternatives with different pay-offs. Users always derive benefits from joining a SNS in terms of increased social contact, therefore this option always scores (1). However, users are also subject to 'tagging' of their photographs and discussions about them if they are not members of SNS (ENISA 2007). If a SNS restricts this form of information uploading without consent, users do not experience the drawbacks from this. When a users joins a SNS and there are no restrictions on collection, processing and dissemination of their information, they will even experience more harm. Social Network Sites do not derive any benefits if users do not join the network, and restricting access to the information of users costs them money. With this pay-off, SNS have no incentive to restrict access to the information, while users have incentives to join the network. Although users seem indifferent between joining or not joining when the access to the information is not restricted, I argue that users are not familiar enough with the negative consequences of joining the network and will therefore choose to join the network.

This simple model suggests that SNS can implement measures that better protect identity-relevant information if these measures would not contradict with creating monetary benefits from the information. Users want the benefits from joining a SNS without the privacy risks. However, many users are unaware of these risks and it it therefore unknown if they would be willing to pay for better privacy protection, thereby creating revenue for SNS from other sources. This should be investigated more thoroughly. Other actors also have an incentive to decrease the privacy risks in SNS because of the negative connotation they generate. All of the partners of the SNS (Beacon, advertisement) have an incentive to prevent this negative connotation of SNS, because this could be coupled to them. Therefore, they want to minimize the negative exposure of SNS related to privacy harms. As these partners have monetary means to contribute to the decrease of privacy risks, this could pose a viable solution for the problem.

The actor analysis of the value chain of SNS shows that many different actors are involved that all have a financial incentive. Social Network Sites currently have no incentives to restrict the access to identity-relevant information. Furthermore, users have incentives to become members of SNS, but get confronted with the negative effects of privacy threats. Governments and privacy regulators have incentives to redefine this imbalance and SNS partners do not benefit from the negative connotation that SNS have gotten. These interests should from the basis for designing a solution that balances the control and access to the identity-relevant information of users more evenly and fairly.

4.4. Specific features of the product

The users of SNS experience the service of the website and derive benefits and detriments from using it. When a user signs up for Facebook, he enters into a contract with the service. The way this contract is constructed influences the control that users have over their identity-relevant information and successively the privacy risks that they are exposed to. In this paragraph, I will draw heavily upon the Facebook case study in appendix A and the CIPPIC Facebook complaint with the Canadian Privacy Officer (Lawson 2008) to describe this contract between the SNS and the user. The CIPPIC complaint is one of the most elaborate descriptions of the workings of SNS and I conducted the Facebook case study to find evidence

for the various interactions between users and a SNS. In addition, I will refer to the work of danah boyd. She is recognized as an expert on the field of interaction between users and online social networks. I will assess this contract between Facebook and the user in terms of fairness, deceptiveness, bargaining power and possibilities to unbundle.

User experience a SNS as an identity-management tool (boyd 2007). They produce their identity online by creating a profile, initiating conversations with others through short messages, adding others to their friends list, placing messages on others public testimonial site, communicating through photos and discuss real-world occurrences on the SNS (boyd and Heer 2006). For many teens, a SNS has become part of everyday life. They get cues on SNS about what to wear, how to act, what is cool and they 'hang out' on SNS (boyd 2006). Through mediated conversations they write themselves into being and this includes the collection, dissemination and processing of their own and others' identity-relevant information.

Writing oneself into being is promoted by the default settings of the SNS. If the default settings show your photo library to anyone in your network, as is the case with Facebook, unwanted dissemination of these photos to people outside your friends list is a fact. To examine which other default settings do not protect users' privacy well, I recorded all the default settings of Facebook in the case study. The only information a user has to provide when signing up are a valid email address, full name, password and date of birth. The email address is needed to validate the user, but full name and date of birth are not necessary to run the SNS and conditioning the use of the service upon providing them is thus in violation of the purpose specification principle. The other default settings align with the three most used services on Facebook: Newsfeed, Beacon and third-party applications. The Newsfeed shows your friends what activities you have performed on the website. Users may opt out of some activities, but by default Facebook Newsfeed disseminates information when you remove profile information, write a wall (public testimonial) post, comment on a note, photo, video or posted item (for sale), post on a discussion board, add a friend, remove relationship status, leave a group or leave a network. Some of this information is considered sensitive under the European Data Protection Directive and should therefore only be disseminated after the data subject has knowingly opted in. Some profile information, as political preferences, sexual interests, relationship status and religious beliefs are considered sensitive and should therefore be provided with explicit opt-in options. Facebook Beacon collects information from your actions on partner websites. Although users can opt out of the disclosure of these actions, Facebook always gets notified about these actions (see the Facebook case study). Facebook mentioned that it discards this information (CA Security Advisor 2007), but why would it waste resources on getting the notifications of these actions if it discards them when users are not Facebook members or have opted out of the feature? Guilherme Roschke (2008) says about the economics of opt-in and opt-out features with respect to third-party application providers:

> *"When you restrict default sharing, processing of information by third parties requires consent by the user. And that requires the user to be informed and asked about the change. When you look at the economics of the situation: the user doesn't know much about how the application works, but the application provider does. In terms of an efficiency perspective, it would cost the user much more effort to get to know how the application works then it would costs the application provider to explain. Putting the burden on the application provider makes more sense. [A]n opt-in system for data collection and*

processing seems much more economic. "

Information crossing different websites and spheres could severely harm users, especially if this happens without their knowledge or consent. The default settings for third-party applications are even more problematic, as they leave the user no choice but to default to giving access to all their information and all the information they can see from their friends (Felt 2008b). Third-party developers get access to all your information and the information you can see from your friends when you add third-party applications. This violates the principle of least authority, as these applications do not need access to all this information to perform their function well.

The conditioning of services with giving up certain information is unfair when this information is not primarily needed for the function. This is why the Fair Information Practices limit the collection of data to the specific purpose for which the data processor needs the data (Gellman 2008). PIPEDA, the Canadian Privacy Act, states that "an organization shall not, as a condition of the supply of a product or service, require an individual to consent to the collection, use, or disclosure of information beyond that required to fulfill the explicitly specified and legitimate purposes" (Lawson 2008). On several instances, CIPPIC found that these principles were violated. A user can not opt out of the advertisements that use their profile information (Social Ads), while these advertisements are not necessary for the purpose for which the information was collected, namely social networking. The conditioning of the use of the third-party applications is also considered in violation of PIPEDA and CIPPIC mentions that Facebook should get express consent for disseminating such sensitive information. Finally, the user's email address and date of birth are used for initial verification, but Facebook does not specify for which purposes they are used after people have signed up. Thereby they condition the service with providing information that is not necessary for the purpose of the SNS.

Furthermore, the behavior of users is governed by SNS policies, such as the privacy policy, terms of use and other policies. In the case of Lori Drew (see paragraph 8.2) the violation of MySpace's terms of use even led to an indictment. By signing up to the SNS, the user agrees to have read these policies. However, the broad consent that users give by agreeing with these policies, is not sufficient for specific processing of sensitive data (Lawson 2008). Under PIPEDA, organizations should make reasonable effort to inform users about the practices related to the management of personal information, but on two occasions Facebook provided developers with more information about which information they can obtain from users than they did users themselves (Lawson 2008).

Although users experience SNS as identity-management tools that help them write themselves into being, the default settings create a wide array of options for harmful information sharing. These default settings are deceptive and unfair with regard to conditioning the use of the service upon the provision of (sensitive) identity-relevant information. Furthermore, SNS are not upfront about the usage of the information they collect and even deceive users by not being transparent.

4.5. Financial flow and flow of services

Social Network Sites are businesses that make money from the identity-relevant information that users upload to their profiles. As mentioned in paragraph 4.3, their financial

incentives lead them to apply means that contradict with the users' incentives to have control over their identity-relevant information.

The flow of money to and from Social Network Sites is very covert and hard to analyze. Hoegg et al. (2006) mention that Web 2.0 services like SNS are based on the principles of the "free economy", which means that users do not have to pay to use the service. Social Network Sites also experience network effects, because the value of a SNS depends on how many members it has and whether friends of potential users are members of the website. However, SNS do incur costs and these costs should be offset by revenues to create a sustainable business model. Although users do not pay for the usage of most SNS, the revenue generated from their identity-relevant information should make up for the costs the SNS owner has. This incentive to make money from other people's information, leads to the privacy risks as defined in this thesis. The ratio of revenue versus costs is yet unclear. MySpace was in 2006 the second-most viewed website, but had a revenue of 3% of Google's expected revenue (Hoegg et al. 2006). Social Network Sites are clearly having difficulties with monetizing on their success.

However, the analysis of the value chain in paragraph 4.3 and articles in newspapers and journals give an idea of how money and services flow through the value chain of Social Network Sites. ENISA (2007) mentions that "the commercial success of an [sic] SNS depends heavily on the number of users it attracts[.] [T]here is pressure on SNS providers to encourage design and behavior which increase the number of users and their connections." The European agency also mentions that the high monetary value of Social Network Sites suggests that the identity-relevant information that SNS collect is being used to considerable financial gain. None of the Social Network Sites has gone public yet, so yearly reports are not available. However, a few SNS are acquired by large companies. Bebo, which is the largest SNS in the UK and has a big market share in the USA, Ireland, New Zealand, Australia and Canada, was acquired by America Online (AOL) in 2008 for $850 million (McCarthy 2008). Other indications of the monetary value of Social Network Sites are the filing for an Initial Public Offering (IPO) by SNS Classmates.com to raise $125 million (Malik 2007) and News Corporation's acquisition of MySpace for $580 million. Social Network Sites that have not been acquired yet, such as Facebook, Friendster and LinkedIn, obtained significant capital from private equity firms.

Other services that SNS offer are also free of charge. As mentioned before, some SNS offer options to add third-party applications to your profile that offer additional services. These applications provide services as showing your friend's location on a map or calculating a matching love score between you and any contact (Felt 2008b). The information flow between these services works as described in paragraph 8.3, the example of sharing and selling of data to third parties. Users request an application and the third-party developer consequently request the identity-relevant information of the user from the SNS. This information is manipulated by the third-party application and fed back to the SNS for placement on the user's profile and dissemination to his contacts. There is no information about financial flows between SNS and third party application developers.

As described in paragraphs 4.3 and 8.4, a similar information flow takes place between SNS and advertising networks. These networks collect information about user's behavior and preferences from SNS to create profiles of users (Chester 2007). They use this information to build profiles of users and predict future preferences and purchases. With this information,

advertising networks display targeted ads to users of SNS. For example on their own profile or when they are viewing a profile of a contact. As mentioned in paragraph 4.3, users can also be profiled on the basis of information from their social network or on the basis of information that SNS obtain in the future. Although it is unclear if and how this happens, the implications for users are clear. Even if users do not provide harmful information on their profile, this information can be inferred from their network or from future information. As the user has no control over both processes, the results could be unwanted. The economic model in which a marketplace brings a buyer and a seller together is called a 'two-sided market'. Social Network Sites bring the seller of identity-relevant information (the user) and the buyer (advertisers) together. In the recommendations is paragraph 10.2 I elaborate more upon the relevance of this economic model for the analysis of Social Network Sites.

4.6. Conclusions

Social Network Sites have complex value chains in which the users are both consumers and producers of identity-relevant information. Many other actors are involved, that provide additional content or advertisement. There are many laws and regulations that govern SNS, however they seem not to be effective against the privacy risks that are identified in this thesis, because those are still occurring on a large scale. The privacy paradox of wanting to disclose identity-relevant information to a group of friends without the information disseminating outside that social sphere creates additional tensions for restricting the access to identity-relevant information in Social Network Sites.

5. Classification of threats

The adapted[37] privacy taxonomy of Solove (2006) provides a good starting point for classifying the privacy threats into information collection, information dissemination and information processing. But how to determine when information is being collected or disseminated? This depends on the perspective of the subject: one's collection of information is the dissemination of information by another. A consistent and satisfying perspective on this issue is necessary for a useful classification. Solove's (2006) taxonomy focuses on activities that involve harm to individuals. The data subject is the starting point in his taxonomy. Information from the data subject is collected, processed and disseminated as shown in the illustration below. This means that the one who executes the harmful activities is always the reference point. The data collector collects, processes and disseminates the information of the data subject. This data collector is also the point of reference for the classification of the various harmful activities provided here.

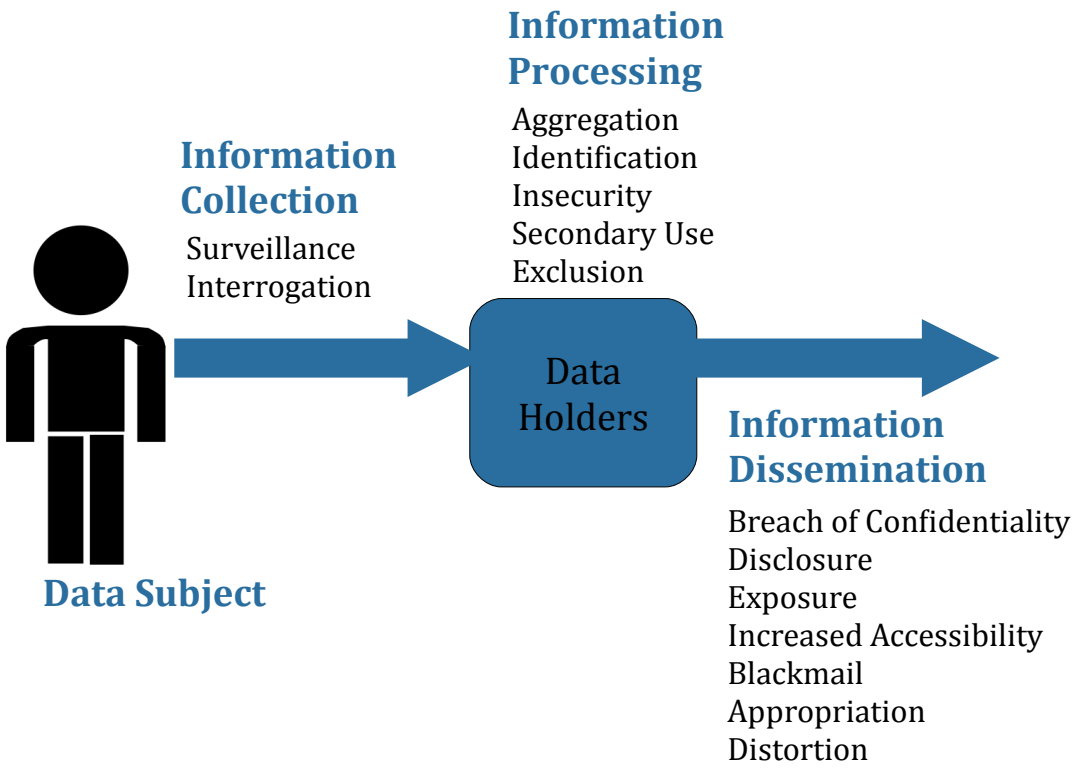


Illustration 7: Adapted version of the Privacy taxonomy of Solove (2006)

Solove's adapted taxonomy is depicted in illustration 7. In paragraph 3.2 I argued that 'invasions' was left out of the taxonomy, because it is normative and the other categories descriptive. In the following paragraphs I will apply the three categories of Solove's adapted taxonomy to Social Network Sites.

5.1. Information collection

Social Network Sites collect information from their users. Solove (2006) writes that information collection creates disruption, based on the process of data gathering. SNS gather

[37] Refer to paragraph 3.2 on why I leave 'invasions' out of the classification.

data when users sign up for the service, when users self-disclose information by uploading it to their profile, when SNS track users' behavior on the websites and when SNS gather information from other sources (third parties). Solove (2006) recognizes surveillance and interrogation as forms of information collection. Although I do believe that the four forms of data gathering above contain aspects of surveillance, I argue that interrogation, as defined in the taxonomy, is less applicable to Social Network Sites.

Interrogation is the pressuring of individuals by the government to divulge information, to compel them to 'testify against themselves'. This type of information collection is not present in SNS. This does not mean that people can not testify against themselves when governments use the information on SNS. In a recent blog (Riphagen 2008f) I gave examples of how prosecutors and lawyers use the information that users themselves post on SNS as evidence. For example, the Fifth Amendment in the Constitution of the United States prevents the government from questioning about one's political associations (Solove 2008). However, Gross et al. (2005) found that more than 50% of the population of Facebook users they surveyed, provided their 'political preferences', for friends to see and download in the public. Users freely give up a right that is protected by the Fifth Amendment. Information collection from SNS by prosecutors and lawyers is still developing and I expect this to become a more common practice.

As mentioned above, the first time that users provide information to a Social Network Site, and the SNS effectively collects information, is during the signup process. Refer to the Facebook case study in appendix A for more information on what the signup process for Facebook looks like. During the signup process, users have to give up their full name, a valid email address (for the confirmation email), their birthday and a password to access the website. From a software perspective, a valid email address (for confirmation and identification of the user) and password would be sufficient to provide access to a website protected with a login. The provision of a full name (referential information) and a birthday (can be used to reverse engineer a SSN) is problematic from a viewpoint of restricting access to identity-relevant information. Furthermore, Facebook's Terms of Use (Facebook 2007b) mentions that a user agrees to “provide accurate, current and complete information about [himself] as may be prompted by any registration forms on the Site”. Optional information that users can fill out during the process is High School and year; College / University and year; company and year; town or city where you live and country of residence. Recall the example from paragraph 2.2 on how identity thieves can reverse engineer SSN by using date of birth and state of birth, as both are connected with the place and the date where the SSN was obtained. Furthermore, controversy arose concerning Facebook's contact importer, which is also part of the signup process (EPIC 2008). This web service allows new users to invite all their contacts from any email service they use or address book they have on their computer. After a user gives his logins and passwords, Facebook logs into these accounts, and collects all the contacts. In a next screen, the user is asked to invite all these contacts to his network, with all the obtained contacts pre-selected. If contacts do not want to be added, the user should opt-out of their selection. Facebook promises not to retain user's passwords and logins. However, Facebook does not explain what happens with the collected email addresses. These are valuable to email harvesters, because they are known contacts of a real person and thus 'live' (EPIC 2008).

After, the signup process, users can add information to their profile. This problem of self-

disclosure is one of the most complex in identifying privacy harms from uploading identity-relevant information to SNS, because information is being published at the initiative of the users and based on their consent (International Working Group on Data Protection in Telecommunications 2008). Because the users upload the information themselves, they are responsible for the initial disclosure[38] of the information. Acquisti and Gross (2005 and 2006) have studied the amount of information that users upload to SNS and have contrasted this with their stated attitude towards online privacy in order to explain the large amount of information that users self-disclose. Recall Susan Barnes's (2006) privacy paradox, as mentioned in chapter one: users want to disclose identity-relevant information to a SNS, but are shocked when this information turns against them. Gross et al. (2005) find, after analyzing more than 4000 Facebook profiles, that more than half of the profiles contain profile images, birthdays, home town and address, AIM screenname, high school, dating interests, relationship status and partner, political preferences, interests and favorite music, books and movies. As explanations, the researchers identified myopic privacy attitudes, peer pressure and herding behavior, the design of the interface with permeable default settings and signaling. Signaling means that users perceive a higher benefit from selectively revealing data to strangers than the costs of privacy invasions. They support this by reporting that single male users tend to post their phone numbers in a higher frequency than female single users or non-single users. Single male users expect to benefit from this, and do not take into account possible privacy threats, such as stalking or being added to a marketing list with phone numbers. ENISA (2007) specifically identifies the risk of stalking through the use of information retrieved from SNS, such as location information and (digital) contact information.

Social Network Sites also collect information on users, while these are surfing the website. This form of covert surveillance (Solove 2006) allows SNS such as Facebook to disseminate information about their users behavior to other users. For example, user X has just wrote a message on user Y's public profile (Facebook's Wall or Hyves' Krabbels) saying Z. I call this internal information collection and it suffers from the same problems that Solove (2006) identifies regarding covert surveillance. It creates an architectural problem[39], or as Solove describes in his book 'The Digital Person' (2004), “an architecture that structures power, a regulatory framework that governs how information is disseminated, collected and networked”. Social Network Sites have full control over which user’s information is sent to whom. This is called informational inequality (Van den Hoven 2007). By collecting information about one person, one gets power over this person, creating a skewed information balance. If SNS know which objects for sale at the marketplace you have looked at during your visits, they are able to offer you exactly those products that you were looking for in targeted advertising. The problem becomes more obvious when SNS are able to collect your bidding prices and are able to anticipate on this. This could lead to unfair harmful activities, such as asking a higher price than previously, because the bidding behavior of a user is known. Van den Hoven (2007) calls this information-based harm, because the harm is made possible with the obtained information.

---

[38] I speak of 'initial disclosure' to identify the uploading of identity-relevant information by the users to the SNS. Further dissemination of the information, for example via Hyves's Buzz to friends in the friends list, is under control of the SNS and therefore I argue that the SNS is accountable for this dissemination.

[39] For a more elaborate discussion of the architecture of control and vulnerability, see the glossary in appendix I

External information collection, the collection of data from sources other than the SNS, is another form of information collection that leads to problems with respect to identity-relevant information. A recent example of external information collection is Facebook's Beacon program, as mentioned in paragraph 3.5. As this example is deconstructed and analyzed in that paragraph, I will not deal with it here again. However, the feelings that users of Facebook voiced regarding Beacon are similar to the harms that Solove (2006) identifies as consequences of surveillance. Acquisti and Gross (2006) mention that 67% of the Facebook users they surveyed mistakenly believes that Facebook does not collect information about them from other sources. People expressed a feeling of discomfort and even anxiety, as all of their relatives and others were notified of their purchases. It is not hard to imagine that you would would be less likely to order certain books for example 'Mein Kampf', if all the people in your friends list are notified of your new purchase. People are represented by digital profiles on Social Network Sites, and people make decisions based on these profiles. The surveillance of people is a good example of a purpose where the information was used, but not intended for. Solove (2006) mentions that surveillance could adversely impact freedom, creativity and self-development, because it leads to self-censorship and inhibition. This is commonly called the 'chilling effect' and Solove (2006) described it as "alter[ing] the way people engage in their activities." In this respect, it restricts the user to engage in identity formation as he wishes, he is pressured from undertaking activities that deviate from social norms and values. This is Van den Hoven's (2007) harm of restricting moral identity and moral autonomy.

The privacy problems from information collection by SNS are imminent. I will classify privacy harms as information collection and group them into the four forms of information collection: signup process, self-disclosure, internal information collection and external information collection. In the next paragraph I will deal with what happens after the information is collected by SNS, the processing of information.

5.2. Information processing

Social Network Sites store the collected information in databases containing users' profiles and process the data for use in their business functions. It is very difficult to obtain information about the information processing and business functions that SNS execute, but it is possible to reverse engineer these activities by examining the public information about SNS, such as newspaper articles, journal articles and investigating the practices of SNS by means of a case study. Solove (2006) defines information processing as the use, storage, and manipulation of the data that has been collected, and he recognizes the following forms of information processing: aggregation, identification, insecurity, secondary use and exclusion. He also describes the processing of information as "connecting data together and linking it to people to whom it pertains" (Solove, supra).

Users create profiles on Social Network Sites that contain various types of information: their personal preferences, name, address and phone numbers, private pictures and videos, and links to friends. Where the photos of your vacation were formerly stored at your house and your links to friends only known by a small group of people; now this information is aggregated[40] and can be found at one central place, your profile. Surprisingly, Acquisti and Gross (2006) find that 70% of the Facebook population they surveyed, does not belief that

[40] For a more elaborate discussion about data aggregation, see the glossary in appendix I.

Facebook combines the information from other sources into their profile. This profile creates a whole that is bigger than its parts, and reveals facts about a person that were previously unknown. The borders of the social spheres of justice are crossed and information is used beyond the expectation of the data subject. This is problematic with respect to the use limitation principle in the Fair Information Practices, which states that information can only be used for the use for which it was initially collected (Gellman 2008). The SNS profile becomes what Solove (2004) calls a 'digital person' and this creates moral problems with respect to informational inequality and the power others have over the profile subject[41]. As this digital person becomes more and more a proxy for our 'real' person, the one who controls the information in the profile becomes more powerful. Distortion and incompleteness of a profile can create severe harm to users. Think of a racial joke made by a stand-up comedian in a private setting. What if we do not know that this person is a stand-up comedian and that the setting is private, would we conclude he is a racist based on the information we have about him?

However, the picture becomes even more complicated if we do not only recognize the person who made the racial joke, but also identify him by first and last name. We attached the information we knew to a real person. This is called identification, and according to Solove (2006) it resembles aggregation because it combines different pieces of information, in this case the racial joke and the joker's identity. Solove's digital person (2004) is now connected to a real-life person. Facebook requires users to sign up with their real name, stating in their terms of use (Facebook 2007b):"[Y]ou agree to (a) provide accurate, current and complete information about you as may be prompted by any registration forms on the Site ("Registration Data")". For SNS, using one's real name creates two potential outcomes. The first is that anonymous speech is made impossible, which chills free speech. In the United States, a woman has been indicted for violating the terms of use of MySpace, because she created a fictitious profile under a fake name (2008a). In other words, she did not provide her real identity. Although this case essentially deals with the suicide of Megan Meier, a case I will discuss in chapter 8, prohibiting fake identities on SNS could lead to severe harms and chilling of free speech. Whistle-blowers hide their identities most of the time, because they fear repercussions. People who hide themselves from abusive spouses take on fake identities online. Revealing their real identity would lead to serious information-based harm for these groups of people.

The other issue with identification in SNS is that aggregation and identification often go hand in hand on SNS. Consider Dutch SNS Hyves' Buzz, a feature akin to Facebook's Newsfeed, which is used by Facebook's Beacon as described in paragraph 3.5. The Buzz shows you what your friends have been doing on Hyves lately. It aggregates information about your actions on different areas of the website, such as commenting on a discussion, adding a picture or tagging ('gespot') someone else in a picture. Then identification takes place, your are connected with the aggregated data and the information of your actions is disclosed to your friends, under your name. First, this ties your name to different activities performed in different spheres (Van den Hoven's informational injustice) and attaches informational baggage to your personality, because your are connected with activities you have done in the past. This limits self-

[41] Here, I mention the 'profile subject', because it is legally very unclear who owns the information on the profile. In various court cases, ownership of profile information was contested, based on copyright liability of intellectual property infringement.

development, as you and all your friends are confronted with for example conflicting opinions you have voiced in different web fora. The comprehensive picture that the Buzz forms of you prevents you from presenting your own identity, and thus limits moral autonomy and moral identity as defined by Van den Hoven (2007). Furthermore, the Buzz stores all the information it has disclosed about you, and other users are able to go back and forth in time through your Buzz. On Facebook, I was even able to see the activities on users' Newsfeed from before I added them to my profile.

Aggregation of information in Social Network Sites can also lead to future identification. The topic of inference with future information is touched upon in paragraph 4.2 and 4.5. The main issue with future inference is that although a user might have no identity-relevant information that could harm him right now, SNS or advertisers might collect information based on your network that can harm you. For example, based on social network information, users could get targeted ads or would have to pay higher insurance premiums. This is a problem because decisions are made based upon information about the user of which he has no control.

Social Network Sites store the profiles of users in databases on large web servers. These servers are the front-end of the SNS and users can access them via the Internet. This makes the servers at the same time very vulnerable to attacks from hackers or cyber criminals. In one of my upcoming posts on the Privacy in Social Network Sites blog (Riphagen 2008g), I will examine the case of a data breach at Facebook in May 2008. In a letter to the Attorney General of Maryland, USA, Facebook explains that “a temporary code glitch caused the driver's license image of some Facebook members to be available to visitors to their Pages for approximately two hours”. The State of Maryland requires data breaches to be notified to the Attorney General. In the letter, Facebook also mentions that only two Maryland residents are affected by the data breach. In personal communications with Facebook, Christopher Soghoian from CNET found out that Facebook thinks that 100 users were affected by this breach (Riphagen 2008g). Some American driver's licenses (varying per state) contain the owner's Social Security Number, and by obtaining this information via Facebook, identity theft is easily possible. This and other data breaches lead to users’ insecurity regarding how well their identity-relevant information is protected on Social Network Sites. Solove (2006) states that the harm derived from future disclosure that could harm users or delayed repercussions because of data breaches leads to feelings of insecurity for these users. I call these future information-based harm. This is why the OECD (Gellman 2008) enacted the principle of security safeguards. This guideline states that data processing agencies should protect data “by reasonable security safeguards against such risks as loss or unauthorized access, destruction, use, modification or disclosure of data” (Gellman 2008).

Two other OECD guidelines, the purpose specification principle and the use limitation principle (Gellman supra), limit the possibilities of secondary use of the collected data. The purpose specification principle states that the reasons why the data is collected should be specified not later than at the time of data collection and the use limitation principle restrains usage of the collected data to these purposes. According to Solove (2006), uncertainty of the secondary use of identity-relevant information leads to a sense of powerlessness and vulnerability. The user can not foresee and therefore cannot consent to the use of information for any other purpose than the purpose he was notified of. Information on SNS could travel multiple social spheres. For example, your sexual preferences become known to fellow

students instead of only friends, thereby unexpectedly harming you. Social Network Sites in general leave the option open for using the uploaded information for other purposes, albeit anonymized[42]. Facebook writes in its privacy policy (Facebook 2007a) that "Facebook may use information in your profile without identifying you as an individual to third parties. We do this for purposes such as aggregating how many people in a network like a band or movie and personalizing advertisements and promotions so that we can provide you Facebook. We believe this benefits you." The phrase "we believe this benefits you", is fitting, because it indicates that Facebook judges which secondary use of the information would benefit the user. It is even more staggering to note that Acquisti and Gross (2006) find that 56% of their surveyed Facebook population does not belief that Facebook shares information with third parties. In this respect, the user has given Facebook control over his information and this creates an architectural problem as described by Solove (2004): one party has more power over the other because its possesses information about that party. Van den Hoven (2007) calls this 'information inequality' and says it could harm users because the more powerful party gets a better, but unfair deal. What can users do if they do not agree with the secondary use that Facebook uses their information for? Nothing, because they do not have any control over the information anymore.

This leads to the last form of information processing that Solove (2006) identifies: exclusion. Facebook's privacy policy (2007a) also mentions that "Facebook may also collect information about you from other sources, such as newspapers, blogs, instant messaging services, and other users of the Facebook service through the operation of the service (e.g., photo tags) in order to provide you with more useful information and a more personalized experience." By signing up to Facebook, you agree with this privacy policy and you can expect that Facebook collects information about you from other sources, as happens with the Beacon program. But what if this information is incorrect? For example, Beacon broadcasts that you bought 'Mein Kampf' instead of Obama's book 'Hope', because the Beacon partner website has made a mistake in coding the different book titles and confused them? If you opted out of Beacon, at least your friends do not see this information. But Facebook is still allowed to retain this data and might target you with ads of a political incorrect nature, because of your stated preference. The problem here is that as a user, there is no way to request all the data that Facebook retains about you, so there is also no way of knowing this or correcting it. In my research on SNS, I have never found a function to request all the information the SNS has about you, while this could easily be done by just printing out the records they have of you in their databases. This failure to provide individuals with notice and input about their records leads, according to Solove (2006), to a sense of vulnerability and uncertainty among individuals because it reduces accountability. The OECD has addressed this issue with the individual participation principle (Gellman 2008), which allows a user to obtain information from a data controller, so users can amend faulty information and have real control over their own information. This is especially important because more and more important decisions are made based on this potentially incorrect information. By applying this principle, a user would be able to easily correct the mistaken book title, and a harmful outcome based on informational inequality would not exist.

Information processing is executed internally or between business partners (in a value

[42] For the difficulties arising from describing the privacy problem only as a dichotomy between identifiable and non-identifiable or anonymized information, refer to paragraph 2.2.

chain). Information dissemination, the disclosure of identity-relevant information to others, is described in the next paragraph.

5.3. Information dissemination

The privacy problems that receive the most attention in Social Network Sites stem from the dissemination of identity-relevant information to others. Solove (2006) describes this as "one of the broadest groupings of privacy harms". Facebook's Newsfeed, which shows your contacts the actions you perform on the SNS and its affiliated websites, was answered by a huge outcry from users (EPIC 2008). Apparently, users of SNS do not want others on the SNS to know all the actions they perform online, or want to be able to select a specific target group to disseminate this information to. This is a good example of what Van den Hoven calls 'informational injustice', the dissemination of information to different 'spheres of justice', that harms us because we do not want people in all spheres to know everything about us, but we have no control over to whom this information is disseminated.

The first harmful form of information dissemination that Solove (2006) recognizes is 'breach of confidentiality'. This harm is also recognized by American tort law and applies when there is a fiduciary relationship between the data subject and the disclosing party, in which the data subject trusts that his information will not be disclosed. Such a relationship exists between physicians and patients, but experts do not agree on whether or not this tort applies to SNS (see paragraph 4.1). I argue that in the strict sense that Solove (2006) defines the breach of confidentiality, it does not apply to SNS. There is no fiduciary relationship between Social Network Sites and their users, who primarily upload information so others can see it.

However, disclosure of identity-relevant information about oneself or about others could lead to privacy harms. Even if this information is true, it could harm a person if the information is considered highly offensive to a reasonable person and is not of legitimate concern to the public (Solove 2006). The information could harm us if other people can use it to our disadvantages, such as information about our work schedules, so criminals will know when we are not at home. Also, although Facebook asks us for our political preference (see the Facebook case study), the disclosure of this information could lead to severe harm in countries where political dissidents are prosecuted, such as China. Disclosure of such sensitive information threatens people's security significantly and restricts their moral autonomy. By promoting the upload of this type of information to their websites, SNS increase the risk of these harms. Furthermore, if a person changed his mind on political preference, this will still be recorded and retained by Facebook. A recent visit to Facebook showed that I was still able to view former profile pictures and information of people in my friends list, even from before I added them to my friends list. This increases the probability of what Solove (2006) calls "[being] a prisoner of one's recorded past". This recorded past could contain information from different social spheres, thereby not only inflicting information-based harm, but also informational inequality.

Solove (2006) coins the term "exposure" to describe the dissemination of embarrassing information about others. Recently, I commented on a case involving exposure in the Netherlands (Riphagen 2008h). A Dutch woman was convicted because she wrote on a 'private' page on Hyves that her ex-husband was a pedophile. The judge convicted her for

libel, because she could not prove the information was true and he considered the 'private' page public. Although there is discussion about whether or not the page was public, her action certainly involved exposing attributes of her ex-husband to a wider public with the explicit goal to create harm. This information-based type of harm can strip people of their dignity and give them a strong feeling of shame (Solove 2006). In this case, it is especially harmful because pedophilia is both illegal in the Netherlands and considered as socially abhorrent behavior. The woman disclosed information that could be used to judge and assess her ex-husband's character. Harm becomes especially visible when the exposed information deviates significantly from social norms, impeding a person's ability to participate in society (Solove *supra*). Social Network Sites easily facilitate the exposure of such information. However, experts did not really agree on to what extent the false light tort and the publicity of private facts tort (which covers both exposure and disclosure) address privacy harms in SNS. As mentioned in paragraph 4.1, experts' opinion on the false light tort ranged from addressing the privacy harms in SNS 'a little' to 'not really.' I argue that this is because it is very hard to prove the 'false' in false light when confronted with the amount of information on profiles of SNS and the large variety in truthfulness of this information. The experts did think the publication of private facts tort 'mostly' addressed the privacy harms in SNS.. Solove (2006) mentions that courts also recognize this tort and that exposures have fared better than disclosures. The disclosure of embarrassing information, for example by uploading nude pictures of others to SNS and tagging those, would be a civil offense under the publication of private facts tort, if the court recognizes the harm of exposure.

As previously mentioned, Social Network Sites also store the history of their users, thereby increasing accessibility to the users' history and making them prisoners of the past. A commonly heard statement from harmed users of SNS is that they thought that the information they uploaded was already publicly available, so it could not harm them (Edwards & Brown). Solove (2006) keenly points out that the development of online court records, which were previously only available offline, could enhance the risk of disclosure to a large unwanted audience. For example, the names of victims of abuse may not be protected by law. The increased accessibility of identity-relevant information, through SNS profiles, also increases the risks of the harmful secondary use of the information. An example of such harmful secondary use is the collection of customers preferences, which happened to a group of tourists while they were staying in hotels in Italy (Article 29 Data Protection Working Party 2007). The collected data about length of stay was added to a marketing profile (processed) and disclosed to the marketing department of the hotels. You give information about your length of stay to a hotel for the purpose of reserving a room, not for marketing and profiling purposes.

Others that have access to your SNS profile, could benefit from the increased accessibility of your information by performing malign activities, such as blackmail or appropriation. I consider both activities information-based harms, and not specific to Social Network Sites, because they can be achieved with any type of identity-relevant information. However, the increased accessibility of identity-relevant information on SNS magnifies the likelihood of blackmail, or, as Solove (2006) defines, the threatening to expose personal information in exchange of demands by the blackmailer. The threat of disclosing the information creates a power relationship between the blackmailer and the data subject, making the latter vulnerable. The increased accessibility of information about ourselves (pictures, videos) makes it easier for companies to use our likeness for their own benefit. A

good example is the dissemination of our purchases, alike endorsement, to our friends via Facebook's Beacon. Facebook's term for this type of viral marketing is 'Social Ads'. If we do not consent to this, as was the case in the first versions of Beacon, this is appropriation. According to Solove (2006), the right of publicity tort and the appropriation tort both address the harmful activity of appropriation. The harm of appropriation stems from the fact that an individual is not able to shape his own identity anymore, a loss of liberty as recognized by Van den Hoven (2007) in privacy harms affecting moral autonomy and moral identity. The experts surveyed agreed most that the appropriation tort addresses privacy harms in SNS 'a little' to 'mostly'. As mentioned above, appropriation seems evident in the Facebook Beacon example. It is therefore not surprising that legal scholars have written plenty on how Beacon resembles appropriation. Daniel Solove (2007b) mentions that Facebook uses the likeness of its users to advertise products for their Beacon partners. He also writes that to avoid the appropriation tort, Facebook should ask the user to explicitly consent to the usage of his name or likeness in an advertisement. William McGeveran (2007) writes that the American law "treats advertising uses differently from other uses." Therefore, the broad general consent that Facebook obtains to use its users information, can never suffice the written consent for the purpose of using one's name to advocate trade, as is required by New York state law. In a follow-up post, McGeveran (2007a) dives deeper into the consent problem. Facebook's privacy officer, Chris Kelly, mentioned that users whose information is used in these ads should not object, because they have chosen to publicly identify themselves with the product. McGeveran mentions however that the liking of the product is different from being presented as a "celebrity endorser" for the product. Facebook should, in his opinion, not transfer consent from some sharing of information for some purpose into "consent for all sharing, including sharing placed in a different context that implies endorsement." Solove (2007c) adds to that that if a celebrity is photographed in public drinking a bottle of Coca-Cola, this does not make it legal for Coca Cola to use this photo in advertisements. McGeveran (2007a) concludes that not many cases have been brought against Beacon, because proof of monetary damage is needed. California's statute includes a fixed 750 dollar penalty. He ends with "if this program does violate privacy laws, the fact that you can get away with it is not a satisfying rationale for going ahead".

Distortion is another form of information dissemination recognized by Solove (2006), that impedes upon one's moral identity. As the 'Mein Kampf' example in paragraph 5.2. illustrates, erroneous information about a person can lead to severe misunderstandings and harms. Inaccuracies in information could lead to people having false beliefs about the data subjects, and they influence the way a person is perceived and judged by others. To prevent this, the OECD has devised the principle of 'individual participation', as described in the previous paragraph. By abiding by this guideline, users get access to the information a data processor collects about them and the opportunity to correct faulty information. Distortion of information has a bigger probability of occurrence in SNS, because other people are able to upload information about you. Friends can upload photos and videos of you and tag those, but most SNS also have an area dedicated to writing testimonials about your friends on their profile. Facebook calls this 'The Wall', Friendster 'Testimonials' and Hyves 'Krabbels'. boyd (2006) notices that teenagers use these public spaces to write personal messages to the profile owners, instead of using email. This means they would post identity-relevant information on these public spaces, increasing the likelihood of false personal information in the public space. boyd *(supra*) notes that "teens are taking social interactions between friends into the public sphere for others to witness".

After this broad discussion of how Solove's (2006) privacy taxonomy applies to Social Network Sites, I will turn to how the taxonomy interacts with Van den Hoven's (2007) moral reasons to restrict access to our identity-relevant information and how this combination can be used to classify privacy harms in SNS.

5.4. Merging the privacy taxonomy with moral reasons to protect privacy

The moral reasons to restrict access to our information correlate with the activities that require access to this information and can inflict privacy harms. In table 2 this correlation is depicted, with a few examples from SNS that I will shortly touch upon below.

	Information-based harm	**Information inequality**	**Informational injustice**	**Moral autonomy and identification**
Information collection		Market failure, no bargaining power or unbundling		No implicit consent to collecting data from third parties
Information processing			Combining data from different spheres in a SNS profile	
Information dissemination	Identity theft, reverse engineering of SSN			

Table 2: Correlation between Solove's (2006) Privacy taxonomy and Van den Hoven's (2007) moral reasons to protect privacy

The case of identity theft by reverse engineering a Social Security Number is explained in paragraph 2.2. This involves dissemination of information from the SNS to the identity thief and severe financial harm. Identity theft is only possible with the obtained information and therefore information-based harm. SNS have more information about exactly how the uploaded information of users will be processed and disseminated, and users can not unbundle the services that SNS offer, which leads to a market failure with respect to bargaining about the user's information. He can upload, for example, his name, but cannot ex ante avoid that his name is connected with pictures that others upload of him. Because the SNS has more information about him than the user, there is an information inequality. This leads to harm, because the user is presented a take-it or leave-it option. There is no possibility for negotiation about the collection of his information. Information collected by third parties that is displayed on SNS limit the user in forming his own identity on a SNS, especially when he has not given his consent. For example, you might not want everyone in your friends list to receive information from which objects you have bought on eBay. This could also lead to information being disseminated to people from different social spheres, as you might not want your co-workers to see which books you order, but do want the people from you weekly book club to see this. However, a user is not given this option. This is called informational injustice, caused by the information processing of the Social Network Site.

6. Privacy threats: acquiring input from privacy and Internet experts

To obtain a broad and valid list of the privacy threats for users of Social Network Sites, I surveyed experts on privacy and the Internet. I was able to contact many of the participants via the the Electronic Privacy Information Center (EPIC), my host in the United States of America. This chapter describes the methodology of the survey and following analysis, from drafting the survey to interpreting the results. For the results of the survey, refer to the next chapter.

Initially, I asked experts to identify privacy harms for users of Social Network Sites. During the course of research, it became clear that the activities that inflict damage to the user could not be classified as 'harms' only. Therefore I adopted a risk analysis perspective, distinguishing threats from incidents or materialized threats. Experts were asked to identify threats to the privacy of users of Social Network Sites[43]. The tort laws provide compensation for the damages that users incur because of threats that materialized in incidents. In chapter 8 I will provide specific examples (incidents) of the threats that materialize and describe the damage they inflict.

The topics of privacy in SNS and the ensuing privacy debate are complex enough without statistical analysis. The survey and interpretation would not have been such a success without the help of Scott Cunningham and Maarten Kroesen from the faculty of TPM, and Lillie Coney, Marc Rotenberg en John Verdi from EPIC.

6.1. Goals of the survey

The results of the survey should provide answers to the following question from the first chapter:

> *What privacy incidents do users of SNS face, what is the prioritization of these incidents and how do the threats that cause these harms interact with each other?*

To answer this question, I have devised the following goals for the survey:

- Obtaining a valid and extensive list of privacy threats for users of SNS.
- Insight in the malevolence of the privacy threats, the amount of damage they inflict or negative impact of these threats for users of SNS.
- Insight in how many people are affected by these threats or the probability of occurrence on a large scale of these privacy harms.
- Classification of the privacy threats to understand how they interact and impact users.
- Prioritization of the privacy threats to create an agenda of privacy harms to address for policy makers.

The target group for the survey are American privacy and Internet experts, because they

[43] Because the realizations that privacy harms consisted of multiple concepts occurred during the course of research, in most of the appendices the aggregated term 'harm' is still used.

are knowledgeable on the subject and the network of EPIC provides easier access to these experts. In the next paragraph I will describe the initial methodology and the adjustments I made to it during the process.

6.2. Methodology

In February 2008, I drafted a document detailing the methodology of the survey. This document can be found in appendix D and was reviewed by the graduation committee. Initially, the survey consisted of the following three aspects:

1. Asking experts for their participation and asking them for referrals to other experts for the survey (co-nomination).

2. Surveying the experts in the first round. This included asking the experts individually to identify threats for privacy in Social Network Sites and the individual rating of these threats on 'probability of occurrence on a large scale' and 'negative impact on users'.

3. Feeding back a summary of the data to the experts and asking them to fill out the survey again, considering the data from the first round. Experts would rate all the privacy threats, even those that others identified, on 'probability of occurrence on a large scale' and 'negative impact on users'.

The identification of possible other respondents by the initial experts helps to increase the sample size. This method is called snowball sampling (n.d.), and contributes to identifying respondents with knowledge of privacy and Internet issues. It builds on resources of existing networks, so the starting group is very important. By choosing the staff of the Electronic Privacy Information Center as starting point, I believe I have identified respondents with enough authority to answer questions regarding privacy threats in Social Network Sites and co-nominate other experts. As mentioned in appendix D, the desired sample included 100 experts. The initial list of experts included 20 experts from the network of EPIC and identified from recent publications on privacy in SNS. However, if not all the initial respondents answer, the amount of non-response goes up and the expected amount of referrals goes down.

In the first round experts would be asked for their participation and for general characteristics such as name, email address and expertise. In the second experts were asked to identify privacy threats in the four classifications of Solove (2006) and rate these privacy threats on 'probability of occurrence on a large scale' and 'negative impact on users'. The response from the second round would be analyzed by me and fed back to the experts in an additional round of surveying. In this round, the experts would rate all the privacy threats, also those mentioned by other experts. This results in a long list of privacy threats, rated on the two factors by all experts.

However, after a few tests, I doubted whether a second round of surveying would be possible, because the very busy schedule of the experts resulted in a high non-response. The first round (asking for participation and nomination of other experts) and the second round (individually identifying and rating privacy threats) were combined. Furthermore, EPIC's Lillie Coney suggested asking the experts how the specific American privacy tort laws address the privacy threats that were identified in the survey. I will contrast the privacy threats' ratings

with these results in paragraph 7.3 to assess the consistency of the responses. The rating of the tort laws occurs on a 5-point Likert-type item, rating from 'does not address' (1) to 'significantly addresses' (5). Because not all response levels were anchored with verbal labels, the scale does not fulfill the fourth condition of a Likert item (as mentioned in appendix G), and is therefore a Likert-type item. I argue that the same calculations used for Likert scales may be applied, as is common practice.

However, response to the first mailing of the survey was very low (only two responses), and this required me to adjust the methodology. After consultation of the graduation committee and the staff at EPIC, I made the following adjustments:

- The classification of privacy threats was left out of the survey. This shortened the survey from 12 questions to 3 questions. The survey would take less time to complete, thereby increasing the likelihood of participation and decreasing the non-response. Classification of the privacy threats into Solove's taxonomy would be done *ex post* by me.

- If different respondents named the same privacy threats, they would be combined. As Likert-type items are of an ordinal measurement scale, they may be summed to represent the score of a group and are therefore also known as summative scales (Uebersax 2006).

- The second round of the survey, which included sending a summary of the results of the first round to the respondents in order to obtain a second rating on all privacy threats, was omitted from the process. This would shorten the time to complete the whole survey. Also, the graduation committee and the staff at EPIC expected a very low response on the second round. However, the validity of the results became a bigger point of concern if the results were not validated by a second round of responses.

- To gain better insight into the validity of the results, consensus and dissensus measurement were applied to the results. Critical analysis of the results is necessary to identify the agreement among the experts and distinguish ratings on which the experts agree very much from ratings on which the experts disagree. A higher agreement rate increases the validity of the results.

- The participation in the survey (and therefore accuracy of the measurements) was increased by cold calling respondents and their referrals. With this technique, respondents were called and interviewed by phone. The answers of the respondents were pasted into a digital version of the survey and sent to the respondent for confirmation. On all occasions, the respondent agreed with the representation of his response in the survey results.

- I surveyed respondents at the Computers, Freedom and Privacy 2008 conference to increase the response. I brought attention to the survey during multiple plenary sessions, and surveyed all participants during a bird-of-feathers workshop that I organized. For the presentation and workshop I gave, refer to appendix O.

During the course of research the concepts of 'negative impact on users' and 'probability of occurrence on a large scale' also became more clear. Referring to the risk analysis framework in illustration 3 in paragraph 2.4, experts were asked to identify threats to privacy.

Not all threats materialize into privacy incidents though, and the 'probability of occurrence on a large scale' refers to the amount of SNS users affected by privacy incidents. This probability must not be seen as probability in the mathematic sense of the word, on a scale from 0.0 to 1.0. Experts felt more comfortable with rating probability of a 5-point Likert scale. Their ratings are suitable for a relative approach, which identifies the most important threats. If threats materialize into incidents, they do not damage every user to the same extent. Therefore, the experts rated 'negative impact on users', or the amount of damage done to a user. For the same reasons, this was also measured on a five-point Likert scale.

The final survey can be found in appendix E. The survey was made available digitally in PDF format that could be filled on a computer. It was sent via multiple email lists and available on multiple websites. The survey was also printed out and disseminated among several experts. With these adjustments to the survey and the process of surveying, I conducted the survey over six months from February 2008 until July 2008. In the next paragraph, I will shortly discuss this process of conducting the survey.

6.3. Process of surveying

After the final version of the survey was drafted, I sent the survey per email out to members of the Privacy Coalition, a nonpartisan coalition of more than 30 consumer, civil liberties, educational, family, library, labor, and technology organizations. They have agreed to support the Privacy Pledge, a framework including promotion of the Fair Information Practices, independent enforcement and oversight, promotion of Privacy Enhancing Technologies (PETs) and promotion of federal privacy safeguards (Privacy Coalition n.d.). Also, I contacted all the scholars that I interviewed and asked them to fill out the survey. Because of the low response on this initial round of surveying, I changed the survey and increased the efforts of surveying. At the bi-weekly meeting of the Privacy Coalition at the office of EPIC in Washington, DC, I surveyed attendants with hardcopy surveys. Furthermore, the survey was mentioned as a news-in-brief in EPIC's long-running (15 years) and highly-circulated newsletters, the EPIC Alert. I also surveyed the staff of EPIC and the various EPIC clerks.

Following up on the co-nominations and references, I started building a list of experts together with Lillie Coney from EPIC. To increase the response rate and get more references, I started cold calling respondents, thereby using the network of EPIC as reference. During a 10- to 30-minutes interview via telephone, I asked the respondents to answer the questions on the survey and filled these out in the digital version of the survey. For their confirmation, I sent this document via email to them. All respondents agreed with the characterization of their answers as mentioned in the survey.

In May 2008, the Computers, Freedom and Privacy 2008 conference was held in New Haven, Connecticut and organized by staff from Yale University Law School's Law and Media program. At this conference, I manned a stand at the registration desk, and asked participants to fill out the survey. Furthermore, I held a 5-minute plenary presentation on 'Social Network Sites and Third Parties', in which I brought attention to the survey. I also organized a Birds-of-a-Feather session, a workshop on 'Social Network Sites and Third Parties' and surveyed all the participants. See appendix O for the presentation I gave and the proposal for the bird-of-a-feather session.

Most of the data was collected in digital form, by means of PDF forms and email. The rest

of the surveys were either conducted via telephone or filled out in hardcopy. I aggregated all the data into one spreadsheet, thereby anonymizing the data and abiding by the Fair Information Practices.

6.4. Analysis of data

In appendix F all the results from the survey are ordered in a table. This is an extensive list of 97 privacy threats, identified by experts. The results have been anonymized and each row represents one respondent. The first thing to notice is that two respondents filled out the old version of the survey and mentioned respectively 12 and 9 privacy threats. Furthermore, one respondent did not mention any privacy threats because he is of the opinion that the increased accessibility of identity-relevant information on Social Network Sites does not lead to privacy threats. Except for an incidental non-response, all the respondents rated the identified privacy threats on 'probability of occurrence on a large scale' and 'negative impact on users'. The survey was designed in such a way that respondents only had to fill out the areas of the survey they were familiar with. Therefore, only 20 respondents rated the American tort laws on their applicability to privacy threats in SNS.

The ratings for the tort laws were immediately straightforward to analyze. I imported the data into SPSS and ran descriptive statistics on the data. The outcomes of this analysis can be read in paragraph 4.1. For more information about the consensus on these privacy threats, refer to paragraph 6.5.

The respondents were also asked to prioritize one of the privacy threats they identified, the one that concerned them the most. The prioritization was added to the survey to force the respondents to think more about the threats and how to rate those. Providing raters which these incentives will make ratings more meaningful for them and will increase the probability of normal distributions from the ratings (Allen 2006). This means that the respondents prioritized 29 privacy threats, because one respondent did not prioritize, one did not mention any privacy threats, and two others mentioned more privacy threats and could therefore prioritize more threats. The list of prioritized privacy harms can be found in appendix H. On the basis of the analysis of activities that can harm privacy from Solove (2006) and Van den Hoven's (2007) moral reasons to restrict these activities, I have clustered the prioritized threats into 10 groups, labeled from A to J. The analysis of each question thus considers the aggregate scores on a Likert-type item (Uebersax 2006) obtained from those respondents who chose to answer the question. These 10 aggregated threats were used to classify the full list of 97 privacy threats, which can be found in appendix I. Note that in this list, the prioritized privacy threats have dark-grey cell backgrounds. One of the privacy threats that came from the long list of harms was however not prioritized. Because this harms could not be classified in another category, I created a category for it, 'Posting of information by others', which has a frequency of 5. This brings the total list of privacy threats to the eleven privacy threats mentioned in paragraph 7.2.

6.5. Priority setting and rater agreement measurement

The questions regarding 'probability of occurrence on a large scale' and 'negative impact on users' were framed to determine the priority of addressing the privacy threats with policy measures. Which of the threats is causing us the most damage and should we therefore

concentrate on putting a stop to? Probability of occurrence on a large scale measures the amount of people that are affected by the privacy threat. Negative impact on users measures how damaging a specific privacy threat is for users. For an elaborate discussion of the results of these rating, refer to paragraph 7.3.

Obviously, a high score on 'probability of occurrence' or 'negative impact on users' increases the priority of addressing this privacy threat, but how significant is this importance? That all depends on the ratings and the distribution of the ratings. Because Likert-type items are ratings between ordered-categorical and interval-level ratings (Uebersax 2006), I will use techniques applicable to both measurement scales and interpret the results on a case-by-case basis. The analysis of data on a case-by-case basis to define rater agreement is supported by Bots and Hulshof (1995 and 2000). Their work is relevant for this thesis, because they empirically tested a method to define dissensus with ratings by a group of experts and provided recommendations to come to prioritization of policy issues based on those ratings.

I analyzed rater agreement in the data sets addressing privacy threats by tort laws, and the ratings on 'probability of occurrence on a large scale' and 'negative impact on users' for the various privacy threats. In both cases, multiple ratings by experts are combined to increase accuracy. Rater agreement measurement is used to interpret the reliability of the summarizing descriptive statistic (either the mode, the mean or both). For example, experts have different opinions on how tort laws address privacy threats, and I am mainly concerned with how the summarizing statistic incorporates these different viewpoints.

Bots and Hulshof (2000) define dissensus as a significant difference of the group average from what most users feel. To examine if this is the case, I plotted the descriptive statistics of all ratings, including the mode, mean, range, standard deviations and bar charts of the frequencies of the ratings. These can be found in appendices K and L. I used the frequency tables to identify peaks in the data. As Bots and Hulshof (2000) mention, in most cases dissensus on a rating scale can be conceptualized as similar frequencies on different ratings. In other words, if 5 users score 1 on a 5-point scale and 5 users score 5 on the same scale, the dissensus is large. Therefore it is important to identify peaks in the data, by examining the frequencies table. Peaks are identified by looking at the maximum frequency, or the ratings the the most respondents chose. Other peaks can be defined relative to this maximum frequency, for example as two-thirds of the maximum frequency. Bots and Hulshof (2000) use 75% of the maximum frequency to identify peaks. They also mention that this percentage should be carefully chosen, after experimentation with the data. For the data I collected, I choose two-third or 66% as the threshold to identify other peaks.

Bots and Hulshof (2000) measure the dissensus with the gamma-factor, a custom developed load factor, that uses the peaks and filters out the noise (response under 30% of the maximum frequency) and calculates the standard deviation with this new filtered data. The Excel model for their calculations can be found in appendix J.

I follow a similar approach to Bots and Hulshof (1995 and 2000), in which I use quantitative methods in a customized way to come up with a definition of consensus that matches the collected data. After experimenting with the data I used the following rules to define consensus:

- Either the data has 1 peak and the standard deviation is less than or equal to 1,

- Or the data has less than or equal to 2 peaks and a gamma factor that is less than or equal to 1.

	Intrusion	Public Private	False Light	Appropriation	Right Publicity	Breach Confidentiality
Max freq	7	7	6	8	7	9
Mode	2	4	3	3	3	4
Mean	2.25	3.20	2.95	3.35	2.65	3.05
Difference		-0.70	-0.05	0.35	-0.35	-0.95
Range	3	4	4	4	4	4
2/3 * Max freq	4.67	4.67	4.00	5.33	4.67	6.00
Number of peaks	3	2	3	2	2	1
Standard deviation	0.756	0.548	1.165	1.035	1.082	0.665
Gamma	1.28	1.34	1.34	0.99	1.09	1.8

Table 3 - indicators to analyze consensus for tort laws

The standard deviation shows the dispersion of the data around the mean, standardized in the values of the scale. The meaning of the standard deviation depends on the type of data that is analyzed. Allen (2006) argues that by making the rating of Likert-type items more meaningful for the respondents and providing incentives for the rater to fill out the survey in his own way, the acquired data can be analyzed as a normal distribution. Plotting a normal curve over the different distributions shows that the ratings for the tort laws follow a normal distribution and that the ratings for 'probability of occurrence on a large scale' and 'negative impact on users' do this when the data is one-peaked or more than three-peaked. When the data is two-peaked, it follows a binomial distribution (Allen 2006) and I argue that the gamma factor, with its corrections for noise and peaks, is a better measurement in those cases. Within normal distributions, 68% of the data lays within 1 standard deviation of the mean. So, if the standard deviation is 1, 68% of the data is dispersed 1 around the mean. In other words, if the mean is 3 and the standard deviation is 1, 68% of the data is between 2 and 4. I chose the standard deviation in the first rule to be 1 or less, because this will give a good approximation of how the data is dispersed around the mean. If most respondents chose the mean and the majority of the rest rated one category lower or one category higher, I argue that it is fair and valid to choose the middle category or the mean as summary of their ratings.

I choose a dissensus measurement factor Gamma of less than or equal to 1, because Bots and Hulshof (2000) also use this threshold to identify ratings where there is no disagreement.

The descriptive statistics for the various tort laws are depicted in table 3 above. The torts that have 1 peak and a standard deviation less than or equal to 1 have light-grey cell backgrounds. The torts that have 2 peaks and a gamma of less than or equal to 1 have dark-grey cell backgrounds. Only the breach of confidentiality tort has 1 peak and a standard deviation less then 1 (0.655). The appropriation tort has 2 peaks and a gamma of less than 1 (0.99). From this, I conclude that the experts agree most on the ratings for these torts.

The rest of the rated torts have 2 or 3 peaks. Now I turn to the bar charts of the data to determine the consensus in these cases. All the descriptive statistics for the ratings of the various tort laws can be found in appendix K. The publication of private facts tort and the right to publicity tort both have 2 peaks that are adjacent. Both ratings have a mean that is in between the 2 peaks, and therefore I believe that the consensus of the experts is well summarized by the means of these ratings.

The ratings on the tort of false light have three peaks: on 2, 3 and 5. Because these peaks are not adjacent, it is difficult to conclude that the mean of 2,95 summarizes the ratings of the experts. Therefore, I mentioned in chapter 4.1 that experts do not agree to what extent the tort of false light addresses the privacy threats they identified in SNS.

The tort of intrusion has three adjacent peaks on 1, 2 and 3 and a mean of 2,25. Because the mean is in between these three peaks and the standard deviation is smaller than 1 and the gamma not much greater than 1, I conclude that the experts do agree on the extent to which the intrusion tort addresses the identified privacy threats in Social Network Sites. The standard deviation smaller than 1 means that most rating fall within less than 1 of the mean, which includes the 3 identified peaks.

All the information needed to apply the rule to the ratings of the various privacy threats on 'probability of occurrence' and 'negative impact on users' is shown in table 4 below and can also be found in appendix N.

	A_prob	A_imp	B_prob	B_imp	C_prob	C_imp	D_prob	D_imp	E_prob	E_imp	F_prob	F_imp	G_prob	G_imp	H_prob	H_imp	I_prob	I_imp	J_prob	J_imp	K_prob	K_imp
Max freq	3	3	5	4	7	9	3	6	2	4	3	3	5	7	8	7	3	2	4	4	3	2
Mode	3-4	5	5	2	5	4	3-4	5	5	5	4	5	5	4	5	5	4	2	3	4	5	3
Mean	4.29	4.60	4.25	2.00	4.20	4.00	3.00	4.62	3.33	4.67	4.25	4.50	4.30	3.80	4.10	3.75	3.75	2.00	3.38	3.88	4.20	3.75
Difference		-0.40	-0.75	0.00	-0.80	0.00		-0.38	-1.67	-0.33	0.25	-0.50	-0.70	-0.20	-0.90	-1.25	-0.25	0.00	0.38	-0.12	-0.80	0.75
Range	2	1	3	3	4	2	3	2	4	1	1	2	3	3	3	3	1	2	3	3	2	2
2/3 * Max freq	2.00	2.00	3.33	2.67	4.67	6.00	2.00	4.00	1.33	2.67	2.00	2.00	3.33	4.67	5.33	4.67	2.00	1.33	2.67	2.67	2.00	1.33
Number of peaks	2	2	1	3	3	1	2	1	1	1	1	1	2	1	3	3	1	1	1	1	2	1
Standard deviation	0.756	0.548	1.165	1.035	1.082	0.655	1.069	0.744	1.633	0.516	0.500	1.000	0.949	0.789	0.912	1.118	0.500	0.816	0.916	0.991	1.090	0.957
Gamma	0.45	0.79	0	1.14	0.75	0.54	1.29	0	2.45	0.00	0.00	1.15	0.72	0.5	0.87	1.53	2.00	2.40	1	1	1.57	1.29

Table 4 - indicators to analyze consensus for probability of occurrence and negative impact

The table above shows the mode, mean, difference between the mode and the mean, the range, the value that corresponds with 2/3 of the maximum frequency, the numbers of peaks, the standard deviation and the dissensus measurement factor gamma as defined by Bots and Hulshof (1995 and 2000) for all the privacy threats' rated on 'probability of occurrence on a large scale' and 'negative impact on users'. The ratings that have 1 peak and a standard deviation of less than or equal to 1 have a cell background that is light-grey. The ratings that have less than or equal to 2 peaks and a gamma factor less than or equal to 1 have a dark-grey cell background.

First, I look at the ratings that have 1 peak and a standard deviation of 1 or less. Negative impact on users, for

- Users can not obtain information over or have control of the secondary use of their information (C_Imp);
- Stalkers, predators or bullies make use of the information on and the interfaces of the SNS to harm the user (D_Imp);
- Identity theft, made possible by the identity-relevant information on the SNS (E_imp);
- Government usage of the identity-relevant information users upload to SNS (F_Imp), and
- The sharing or selling of identity-relevant information of users to third parties to create revenue from the information (G_Imp)'

- Displeasure on behalf of the data subject from the feeling of being monitored all the time (I_imp), and
- Damage to reputation of the data subject because of the disclosure of untrue or embarrassing information about him (J_imp)

have both 1 peak and a standard deviation equal to or less than 1, respectively 0,655; 0,744; 0,516; 1 and 0,789; 0.816 and 0,991.

Probability of occurrence on a large scale, for

- Government usage of the identity-relevant information users upload to SNS (F_prob);
- Displeasure on behalf of the data subject from the feeling of being monitored all the time (I_prob), and
- Damage to reputation of the data subject because of the disclosure of untrue or embarrassing information about him (J_prob)

have both 1 peak and a standard deviation equal to or less than 1, respectively 0,500; 0,500 and 0,916.

Next, I distinguish the ratings that have 2 peaks or less and a gamma factor that is less than or equal to 1. Probability of occurrence on a large scale, for

- Online tracking of the activity of users on the SNS and other websites (A_prob);
- Aggregation of the data of users into profiles of users (B_prob);
- Users can not obtain information over or have control of the secondary use of their information (C_prob);
- The sharing or selling of identity-relevant information of users to third parties to create revenue from the information (G_prob),
- The unwanted dissemination of identity-relevant information to others or groups of others, without the consent of the data subject (H_prob)

have 2 peaks or less and a gamma factor less than or equal to 1, respectively 0,95; 0; 0,75; 0,74 and 0,97. Negative impact on users, for online tracking of the activity of users on the SNS and other websites (A_imp) has 2 peaks and both a gamma smaller than 1 (0,79) and a standard deviation smaller than 1 (0,756).

This leaves the following ratings on which respondents do not agree:

- Probability of occurrence on a large scale, for stalkers, predators or bullies make use of the information on and the interfaces of the SNS to harm the user (D_prob), because this data set has 2 peaks, and a standard deviation larger than 1 (1,069) and a gamma factor larger than 1 (1.29). This can be explained because although 6 respondents rate this threat as 'may affect a group of people' (3) and 'will affect a group of people (4), these 2 respondents rate on 'unlikely to affect a significant amount of

people (1) and 'may affect a few people (2), creating a big spread in ratings.

- Probability of occurrence on a large scale, for identity theft, made possible by the identity-relevant information on the SNS (E_prob), because this data set has 1 peak, but both a standard deviation higher than 1 (1,633) and a gamma factor higher than 1 (2,45). This is explained by the ratings of respondents on all five categories, with a peak that has a frequency of 2 on 'likely to affect a significant amount of people' and a frequency of 1 on all the other categories.

- Probability of occurrence on a large scale, for posting of information on the Social Network Site by others, without the consent of the data subject (K_prob), because this data set has 2 peaks and a standard deviation higher than 1 (1,090) and a gamma factor higher than 1 (1,57). This is explained by the fact that 3 respondents scored 'likely to affect a significant amount of people' (4) and 2 respondents scored 'may affect a group of people' (3), thereby creating two peaks and making it difficult to find a summarizing value.

- Negative impact on users, for aggregation of the data of users into profiles of users (B_imp), because this data set has 2 peaks and a standard deviation higher than 1 (1,035) and a gamma factor higher than 1 (1,14). This can be explained because 4 respondents voted 'reversible harm to the users' (2), 3 respondents voted 'harming the user (3) and 1 respondent voted 'causing great damage to the user (4), thereby creating a big spread of answers with low frequencies.

- Negative impact on users, for the unwanted dissemination of identity-relevant information to others or groups of others, without the consent of the data subject (H_imp), because this data set has 2 peaks and a standard deviation higher than 1 (1,118) and a gamma factor higher than 1 (1,53). This is explained because 2 respondents rated the threat as 'harming the user' (3), 1 respondent rated 'damaging the user significantly' (4) and 1 respondent rated 'causing great damage to the user' (5). This creates a big spread of ratings without any significant peaks.

6.6. Conclusions

In this paragraph I have described the process from drafting the goals for the survey and the initial methodology, to the revisions of the survey and methodology, the increased effort in creating response and the data analysis. In this process, I believe I have collected and measured a valid set of data that is very relevant for the identification of privacy threats in Social Network Sites. Finally, I have identified eleven privacy threats and measured them on 'probability of occurrence on a large scale' and 'negative impact on users' and measured the applicability of the American privacy tort laws to these privacy threats. By scrutinizing the data and analyzing rater agreement and prioritization, I was able to assess the validity of the results. This resulted in one less valid rating on tort laws and five less valid ratings on the probability of occurrence and negative impact ratings. In the next chapter, I will translate the numerical findings into conclusions about privacy threats in SNS: what does this data mean for the privacy in Social Network Sites?

7. Privacy threats, incidents and harms for users of SNS

Introduction text about threats for privacy, that materialize (incidents) and create harm for users (damage them). Laws are directed at compensating for the damage incurred.

7.1. Description of respondents

In total, 29 experts participated in the survey. They filled out the survey digitally (a PDF of the survey was emailed to different email lists), were interviewed by phone and their answers translated to the survey or filled out the survey with pen and paper at the office of EPIC or at the CFP 2008 conference. Each of the experts was asked to describe their expertise on privacy issues on Social Network Sites and rate their expertise on a five-point ordered-category item (Uebersax 2006). The experts were asked to rate their expertise as:

1. 'I work on privacy or Internet issues, but not combined'
2. 'The main focus of my work is either privacy or Internet issues'
3. 'The main focus of my work is on privacy issues on the Internet (combined)'
4. 'I have significant experience working on privacy issues on the Internet'
5. 'I am recognized as an expert on the field of privacy issues on the Internet' (5).

As can be seen in illustration 8, most of the experts (73%) focus on privacy and Internet issues, while 28% has significant experience on privacy and Internet issues and 21 % of the experts is recognized as an expert in the field of Internet privacy issues.

From these ratings, I conclude that the respondents are very well qualified to answer the questions of the survey. A majority of the respondents consist of (legal) scholars on privacy and Internet issues. They vary from undergraduate students of computer engineering, to graduate and law school students and law school professors. A significant number of respondents consist of privacy activists and people who work at NGOs that have privacy as a significant field of interest.

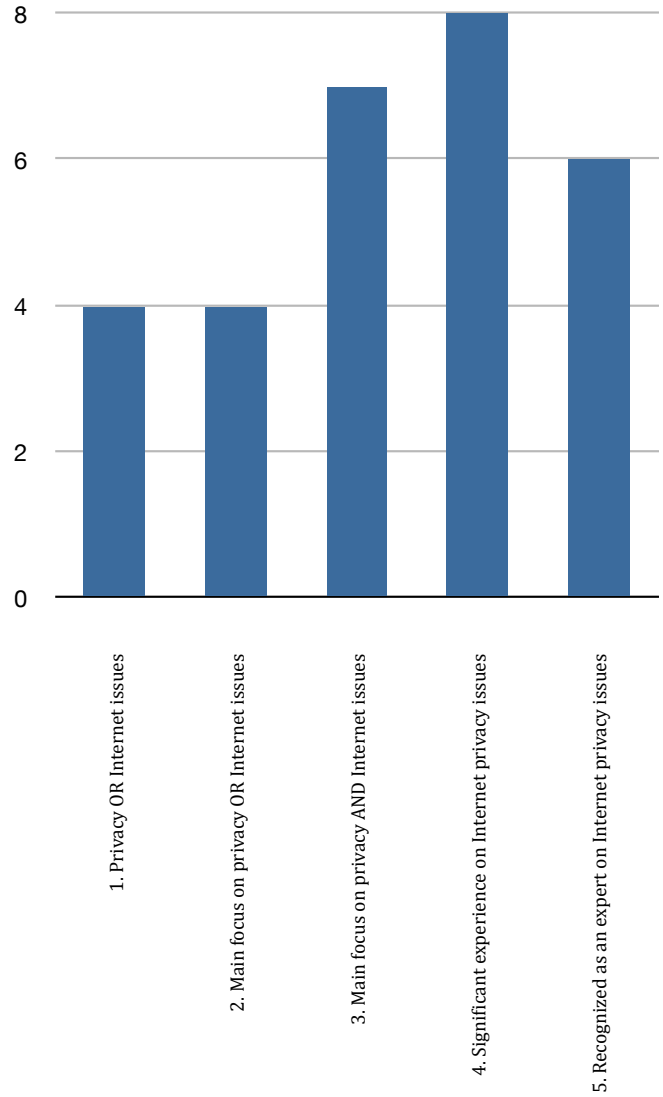


Illustration 8: Ratings on the expertise of the respondents.

7.2. Privacy threats for users of SNS

In total, the experts that filled out the survey identified 97 privacy harms. Most respondents, 24, identified three privacy harms, as was asked in the second version of the survey (totaling 72 privacy harms). One respondent did not identify any privacy harms, because he believes that identity-relevant information on Social Network Sites does not necessarily lead to harm. He did rate the American tort laws on their applicability to privacy harms in Social Network Sites. One of the two respondents who participated in the first version of the survey, mentioned the full 12 privacy threats and the other expert mentioned only 9 of the 12 asked. He noted that he did not understand why the privacy threats were classified in the different categories. Based on his and others' feedback, I changed the survey and left out the classification of privacy harms. Two respondents, who took the new survey, only identified two privacy threats instead of three. This does not influence the validity of the survey, as the basis of the survey is the individual privacy threats and not the number of

threats a respondent mentions. Refer to appendix D for more background on the set-up of the survey and chapter 6 for more information on the methodology of the survey.

In both the first and second version of the survey, the experts were asked to prioritize one out of every three threats for privacy. In the first version of the survey they were asked to prioritize one threats per category. There were four categories: information collection, information processing, information dissemination and invasions. The respondents prioritized 29 out of the 97 privacy threats, a score of 30%, or almost one out of three. There is a bias caused by the experts who did not fill out three privacy threats per category, but did prioritize the ones that they identified.

As mentioned in paragraph 6.4, I aggregated the different privacy threats that were similar to each other. For example, six respondents mentioned the activity of 'identity theft' as harmful. These were grouped into one threat for privacy. Because the Likert-type items are summative (Uebersax 2006), they can be summed and standardized (divided by the frequency) to give an average score for the respondents who identified the same threat. For example, 9 respondents identified the threats to privacy posed by data aggregation and user profiling.

I analyzed and aggregated the privacy threats that the experts prioritized. With every classification (a group of three privacy threats), I asked the experts which of the privacy threats concerned them the most. The analysis of these 29 privacy threats can be found in appendix G. This led to the following list of aggregated privacy threats for users of SNS. The number of times each threats was listed as prioritized are shown in parenthesis.

A. Online tracking of the activity of users on the Social Network Site and other websites (4).

B. Aggregation of user data into specific profiles (5).

C. Users cannot obtain information about or have control of the secondary use of their information (3).

D. Stalkers, predators or bullies make use of the information on and the interfaces of the Social Network Site to damage the user (2).

E. Identity theft, made possible by the accessibility of the identity-relevant information on the Social Network Site (3).

F. Government usage of the identity-relevant information users upload to Social Network Sites (3).

G. The sharing or selling of identity-relevant information of users with third parties to create revenue from that information (4).

H. The unwanted dissemination of identity-relevant information to others or groups of others, without the consent of the data subject (3).

I. Displeasure on behalf of the data subject from the feeling of being monitored all the time (1).

J. Damage to reputation of the data subject because of the disclosure of untrue or embarrassing information about him (1).

Then I categorized the full list of privacy harms, which can be found in appendix I ,by the categories mentioned above. However, a few privacy harms could not be headed under one of the categories mentioned above. Therefore I added another aggregated privacy threat to the list:

K. Posting of a user's information on the Social Network Site by others, without the consent of the data subject (5).

Between brackets I have noted the amount of privacy threats that comprise the aggregated threat. The full list of privacy threats changed the frequency as follows: A (7); B (9); C (14); D (8); E (6); F (4); G (11); H (19); I (5); J (9) and K (5). This increased frequency per privacy threat demonstrates that the aggregated threats were chosen carefully and appropriately summarize the threats to privacy identified by the experts.

7.3. Probability of occurrence and negative impact

These eleven privacy threats (A-K) are the most important threats for privacy on Social Network Sites that experts have identified. As mentioned in paragraph 6.2, not all privacy threats materialize in the same way. Threats that materialize become incidents and not all users are affected by every incident. For example, not all users are affected when the identity of one user is stolen. Furthermore, the damage occurred from these incidents varies per user. Some of the incidents have grave consequences and damage users significantly, such as the suicide of Megan Meier. The amount of users affected by threats materializing into incidents and the amount of damage they do to users are two important measures to identify the graveness of the threat. This research aims to identify the most detrimental ones, so policy makers know which privacy harms to address first. To assess this, I asked the experts to rate the threats they identified on:

- Probability of occurrence on a large scale; in other words, the probability that it will affect a significant portion of SNS users. If an threat materialized into an incident, how many users are affected?

- Negative impact on the users, or how much harm is done to the user by this activity? The damage that an incidents inflicts upon a user differs per incident and per user. This rating measures how much damage a threat can inflict upon a user.

The experts were asked to rate the probability of occurrence on a large scale of each of the threats for privacy that they identified on a five-point Likert item. Only two of the five levels were anchored with verbal labels, so the scale does not fulfill the fourth condition of a Likert item (as mentioned in appendix G), and is therefore a Likert-type item. Uebersax (2006 and 2008) mentions that the same calculations may be applied to Likert-type items as to Likert items. Ex post the following labels can be distinguished:

1. Unlikely to affect a significant amount of people.
2. May affect a few people.
3. May affect a group of people

4. Will affect a group of people
5. Likely to affect a significant amount of people.

How much damage was inflicted on users was measured by asking respondents to rate the privacy threats they identified on another five-point Likert-type item. The scale has the following labels:

1. Little harm to the user.
2. Reversible harm to the user.
3. Harming the user
4. Damaging the user significantly.
5. Causing great damage to the user.

The various ratings on the Likert-type items are summed for the aggregated privacy threats above and divided by the frequency to get the average probability of occurrence and negative impact as identified by the experts. This is depicted in the table in appendix I. The Liker-type items are suited to identify relative ratings, so it is possible to identify which incidents affect more people than others and which incidents infer more harm to users than others. This leads to scores for all the eleven identified threats for privacy on 'probability of occurrence on a large scale' and 'negative impact on users'. These scores are graphically depicted in an X-Y plot in illustration 9. 'Probability' is placed on the X-axis and 'negative impact' on the Y-axis.

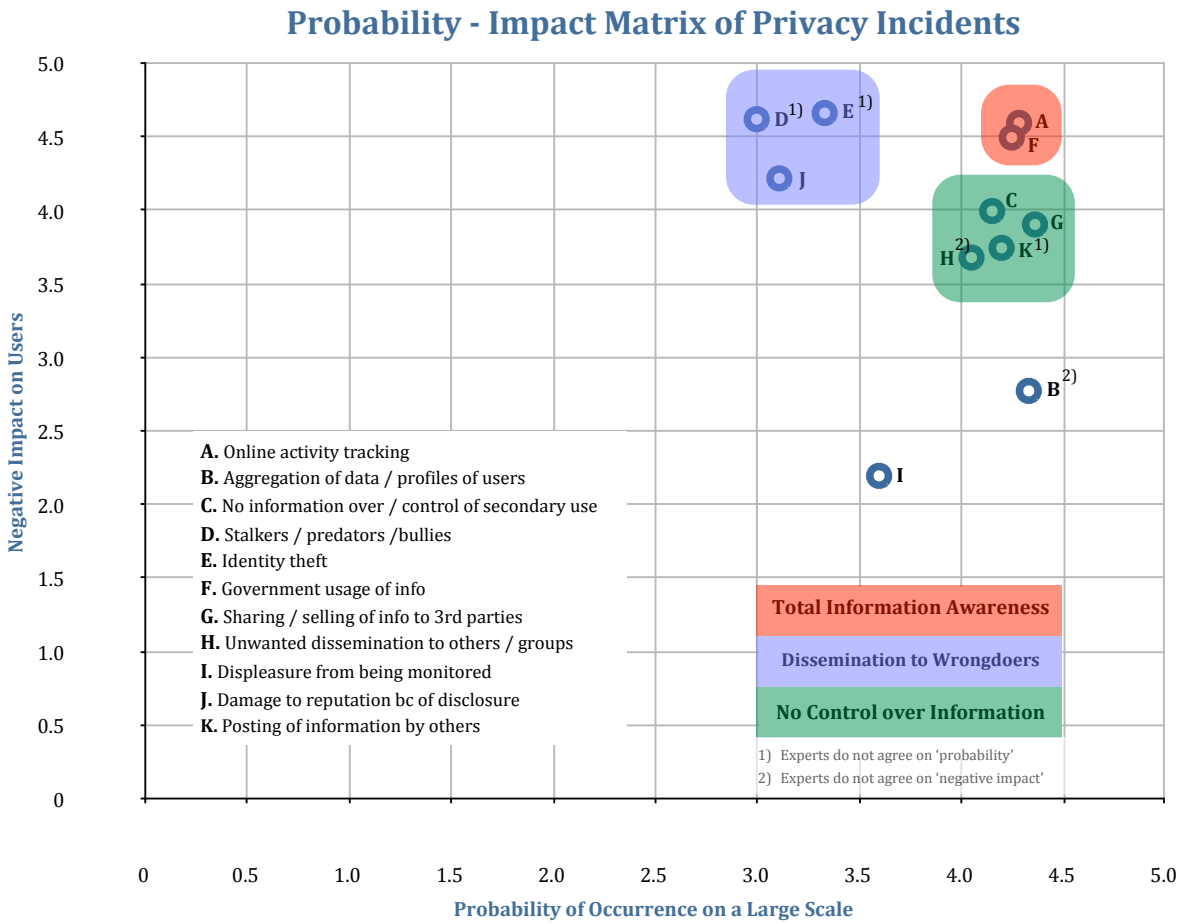


Illustration 9: Probability and Impact scores for the 11 privacy incidents (materialized threats) that experts identified and their merging into three different groups of incidents.

With respect to the distribution of the ratings, note that based on the descriptive frequencies the 'stalkers / predators / bullies' threat (D), could score a little higher on 'probability of occurrence', but that experts do not agree on this. The 'aggregation of data / profiles of users' (B) threat is expected to score a little lower on the 'negative impact' scale, but experts also do not agree on this. Furthermore, experts did not fully agree on the 'probability of occurrence' for 'identity theft' (E) and 'posting of information by others' (K) threats; and on the 'negative impact' of the 'unwanted dissemination to others / groups' (H). This is also depicted in the illustration.

This probability-impact matrix makes it easier to analyze the results in terms of harmfulness for the users and likelihood of a large amount of users being affected by this threat. It is clear that the threats for privacy group together in three clusters. Two privacy threats, 'Aggregation of data / profiles of users' (B) and 'displeasure from being monitored' (I) do not group into a cluster. Those two threats are not dealt with here. They will be analyzed in chapter 8, as these privacy threats could have a significant impact on a user, justifying a more thorough investigation.

Total Information Awareness

The first, most harmful cluster, both in terms of 'probability' and 'impact', is the red cluster in the right-hand top corner. This cluster consists of 'online tracking of the activity of users on the SNS and other websites' and 'government usage of the identity-relevant

information users upload to SNS'. The first activity is information collection, which could lead to informational injustice, as information from different spheres, or in this case, different websites, is being collected. In the ENISA position paper (2007), researchers state that as SNS have become a one-stop communication hub for messaging services, they are able to "gather unprecedented amounts of secondary personal information on their users." The second activity is information dissemination to a specific party, the government. This could lead to an 'architecture of vulnerability' (Solove 2004), as the government has more information about an individual than the individual knows or wants. Van den Hoven (2007) calls this informational inequality. Note that in paragraph 5.1, I mentioned that "interrogation"[44] as defined by Solove (2006) would not be applicable to Social Network Sites and therefore did not elaborate on this category. However, the results from the survey show that this might not be the case, and although information dissemination to the government might be less visible and not be accompanied by direct pressure from the government, experts say that it happens on a large scale and could harm users severely. Information of users is used to let them 'testify against themselves'. The collection of information from many different sources and the use of this information by the government is similar to the Total Information Awareness (TIA) project run by the American government. This project began in 2002 and was created to virtually aggregate data from different public and private sources, and to find patterns in the data that could lead to criminal or terrorist activities (EPIC 2008b). This project analyzed troves of information to detect terrorists (EPIC 2005). Early in the project, Healy (2003) warned that emails and telephone records were also collected. The TIA project is similar to the two privacy harms in that it collects information from various sources, as SNS are doing as well, and disseminates this information to the government for law enforcement purposes. Therefore, I coin the first group of privacy threats (in red) 'Total Information Awareness'. In chapter 8, I will give examples of these threats, how they materialize (incidents) and what kind of damage results from them.

Dissemination to Wrongdoers

The second group of threats for privacy occurs less and therefore affects fewer people, but experts do agree that this group of threats has a very negative impact on the users of Social Network Sites and causes severe damage to them. This group is a little left of the top-right corner and the privacy threats are grouped in a blue area. The threats to privacy belonging to this group are 'Stalkers / predators / bullies' (D), 'Identity theft' (E) and 'Damage to reputation because of disclosure' (J). These privacy threats all have to do with the dissemination of sensitive identity-relevant information to people who perform malign activities with this information. In the case of stalkers, it is probably location information; in the case of identity theft, it is information that is required to request credit; and in the case of reputation harm because of disclosure, it is embarrassing information or information that is considered taboo, such as explicit sexual preferences. ENISA (2007) states that Social Network Sites have increased the vulnerability from cyberbullying because communication can be targeted at specific people, is less visible for teachers and parents, and all tools are available at one website to organize a cyberbulying campaign. Also, note the case of Megan Meier, which describes cyberbullying in detail, in paragraph 8.2. Because the receivers of all these types of information all use the information to do wrong to the data subject, I coin this group 'Dissemination to wrongdoers'. This dissemination harms users, because the information

[44] Solove mentions that "[I]nterrogation is the pressuring of individuals to divulge information." (2006). He applies this concept only to pressuring by the government.

could be used for malign activities (information-based harm) or the information is disseminated to people in different social spheres (informational injustice). In chapter 8, I will give explicit examples of privacy harms from this group.

No Control over Information

The third group of privacy threats has a less negative impact on users and thus damages users less severely, but experts agree that a large scale of SNS users are already subject to these threats. This is a group of threats which experts believe that most users are subject to, but they do not rate the incurred damage as significant as 'dissemination of information to wrongdoers' or 'total information awareness'. This group is a little below the top-right corner and grouped in a green area. The privacy threats in this group are 'No information / No control over secondary use of information' (C), 'Sharing / selling of information to third parties' (G), 'Unwanted dissemination of information to others or to groups' (H) and 'Posting of information by others' (K). While all threats fall under different categories from Solove's (2006) framework, the underlying daamge is clear: the user has no information and no control over how and when his information is collected, processed and disseminated. For example. secondary use falls under information processing. Sharing, selling and dissemination of information under information dissemination. And posting of information by others under information collection. Therefore, I name this group 'No control over information'. The implications of this threat are evident: the data subject cannot control his identity-relevant information and therefore has no way to prevent any threats from materializing into incidents. For example, other users can post pictures of him on the SNS, this information can be shared with people without his consent and even sold to marketing companies. This could harm the users because information could travel different social spheres without his consent (informational injustice), but also because it prohibits him from forming his own identity and therefore restricts him in what Van den Hoven (2007) calls this moral identity and autonomy.

7.4. Tort laws to minimize the damage from privacy threats

Not all threats for privacy in Social Network Sites materialize and damage users in the same way. A risk analysis approach is needed to incorporate these insecurities into the analysis, as described in paragraph 2.4. The experts identified multiple threats for users of Social Network Sites. When these threats materialize, they become incidents that affect a certain amount of people. However, not all of these incidents damage users in the same way. The damage incurred depends on the type of threat and the perception of the user. Laws are directed at preventing the threats from materializing. More specifically, American tort laws create means for people to find compensations against the damages they incurred. Therefore I asked the experts to rate to what extent the tort laws address the threats for privacy and resulting damages they identified. The ratings of the various tort laws are depicted in illustration 10 below.

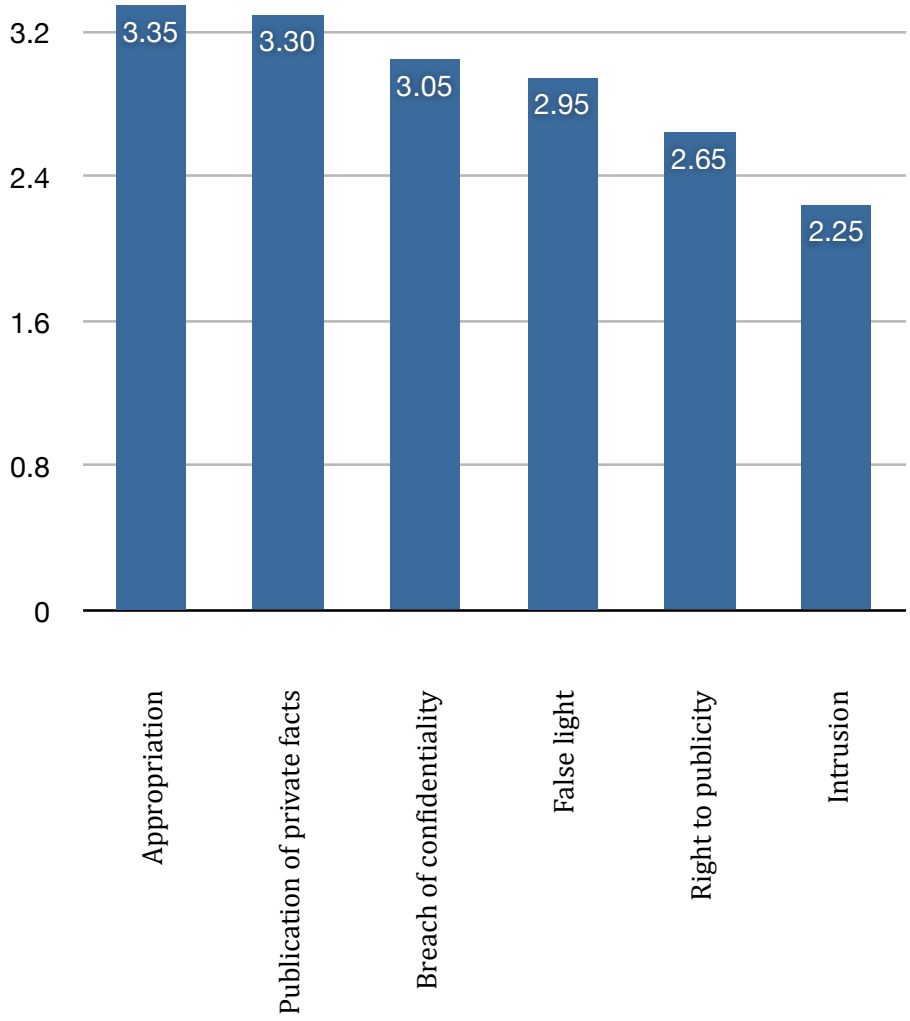


Illustration 10: To what extent do the American tort laws address privacy threats in Social Network Sites and the resulting damages?

Total information awareness

The two problems with Total Information Awareness originate from information collection from sources without user consent and the dissemination of this information to the government for law enforcement purposes.

The collection of information without user consent could be considered illegal under the intrusion tort, the breach of confidentiality tort, the publication of private facts tort and the right to publicity tort. Which tort law applies depends on the type of information collected and damage incurred from the collection of this information. The breach of confidentiality tort, which addresses the privacy threats in SNS **'a little'** according to experts, only applies when the collected information is related to a fiduciary relationship. The question of whether there is a fiduciary relationship also depends on the kind of information that is shared between the parties, leaving us with a circular argument. However, there are plenty of examples in which information of a confidential, if not explicitly fiduciary, nature is shared. For example, multiple sources have confirmed that JP Morgan Chase, an American bank, and Facebook teamed up to promote a Chase credit card (Simon 2006, Cashmore 2006, Ruhami 2006). The actions that you perform on Facebook are coupled to reward points you can earn when using your credit card. To link your actions to the credit card account, Facebook needs access to confidential information that could place them in a fiduciary relationship with you. However, details of this

program are unknown.

According to the experts, the publication of private facts tort addresses privacy threats in SNS **'a little'** to **'mostly**'. This tort imposes liability when (1) one publicly discloses, (2) a matter concerning the private life of another, (3) that would be highly offensive to a reasonable person and (4) that is not of legitimate concern to the public (Bright 2008). Therefore the private information must first be collected and then disseminated. A recent case shows how this tort can protect users against the collection and online dissemination of personal information. The British newspaper, 'News of the World,' posted a video online that showed Max Mosley, the director of the Formula One, performing sexual acts with five prostitutes in what looked like Nazi uniforms (Bayard 2008). Mosley sued the newspaper and England's High Court ruled in his favor. The publication of private facts tort requires that the publication of the information is highly offensive to the person and not of legitimate concern to the public (The Reporters Committee for Freedom of the Press 2003) and the Court considered the sexual activities by Mosley not newsworthy and mentioned that the newspaper had no proof of the Nazi theme of the activities. However, a potential problem with the application of this tort to Social Network Sites is the definition of 'private' information. boyd (2007a) struggles with the question of whether SNS are public or private space, describing SNS as places where people publicly gather and are mediated by technologies. She mentions that "[w]hat it means to be public or private is quickly changing before our eyes and we lack the language, social norms, and structures to handle it." Also, she concludes that most teens rely on 'security through obscurity', or in other words: because the likelihood that their information will be targeted by wrongdoers is not that high, they consider their information safe.

The right to publicity tort, which experts rate as addressing privacy harms in SNS **'a little'** to **'not really'**, says that a person has the exclusive right over the publicity given to his performance. A person thus has the right to control the commercial use of his or her identity. While this tort protects, for example, celebrities against the use of their videos and appearances on other websites that then profit from it, it may be difficult to see how this would protect users of Social Network Sites. Quinn (n.d.) characterizes this right as a property right and states that it can be triggered by any unauthorized use in which the plaintiff is 'identifiable'. The problem with 'unauthorized use' is that users self-disclose information to Social Network Sites, as discussed in paragraph 5.1, and that users agree with the very broad and vague policies of SNS (see paragraph 4.1) for the secondary use of their information. It is therefore not surprising that experts do not think the right to publicity tort protects users of SNS against privacy threats.

Experts think that the intrusion tort does not really address privacy harms in Social Network Sites. This tort also suffers from the problematic dichotomy between public and private space that Solove (2006) calls the 'secrecy paradigm'[45]. Therefore I can conclude that this tort does not address the privacy harms in SNS well.

No specific tort law addresses the dissemination of the information to the government

[45] Solove (2006) mentions that the secrecy paradigm is used to define privacy harms as the result of the publication of private information in a public environment. Solove adds that "[I]f the information is not previously hidden, then no privacy interest is implicated by the collection or dissemination of the information."

for law enforcement purposes. However, Solove (2008) does mention that there are implications regarding the Fifth Amendment to the US Constitution, which protects an American citizen from 'testifying against oneself'. The collection of data in fusion centers (EPIC 2008) by the government has increasingly gained public attention, and the concerns of the experts (and many others) are justified, especially when we take into account the sensitivity of the information posted on Social Network Sites (such as political views).

Dissemination to wrongdoers

A person's information may be used to harm him if it is disseminated to certain parties. Tort laws that address this problem restrict the harmful activities a person can conduct with someone else's identity-relevant information. In the preceding paragraph, I already mentioned that the right to publicity tort does not fully address the dissemination of information to wrongdoers. The publication of private facts tort does a better job, but because publication on SNS is done by users themselves, it is difficult to distinguish who exactly is liable for this dissemination.

Wrongdoers use the information on SNS to harm users in a emotional, financial or reputational way. The tort of false light protects people against the dissemination of information that portrays them incorrectly or puts them in a false light. Experts did not agree on to what extent this law protects users of SNS and Martin (n.d.) mentions that it is one of the most problematic tort laws. False light requires some type of false or misleading information. If the information can be objectively confirmed as false, it is a case of defamation. However, the problems with the false light tort derive from the fact that one has to prove that (1) the publication (2) is made with actual malice, (3) places the plaintiff in false light and (4) would be highly offensive. The disagreement between experts is caused by the proof of the actual malice and the offensiveness of the information. boyd (2007a) mentions that youth use SNS as playgrounds, and therefore "shaming and naming" is common practice. It depends on the case whether or not could we speak of malicious intent and offensiveness.

No control over information

There is a tort law that explicitly addresses the control over one's information. The appropriation tort makes someone liable when he uses the likeness of another for his own benefits. In paragraph 5.3, I discussed how Facebook's Social Ads combines a user's profile picture and name with a product, based on his action on Facebook. McGeveran (2007a and 2007b) and Solove (2007b and 2007d) both argue that this can be addressed by the appropriation tort. But McGeveran (2007a) mentions that the monetary damage is difficult to prove and the benefits of suing Facebook under this tort are low. However, experts think the appropriation tort addresses privacy threats in SNS better than all the other torts they have rated. They rated it as addressing privacy threats in SNS 'a little' to 'mostly'. I argue that if the aforementioned difficulties are resolved, the appropriation tort would be useful in restricting the harmful activities on SNS.

Furthermore, the posting of information by others might be subject to the publication of private facts tort, the right to publicity tort and the false light tort. However, there has not been much discussion on this. The sharing and selling of information to third parties could be considered appropriation, but again, the proof of damage is difficult and the incentive to sue low.

Some tort laws are adequate to prevent harmful activities on Social Network Sites, but in

practice they suffer from many drawbacks. Lee Tien (2008) says about these drawbacks:

> *"The privacy torts don't cover the different services in the Internet Age very well. For one, it is very difficult to meet the requirements for the tort: [..] when does harm actually take place [..]? [..] [Furthermore,] most dissemination of information that users experience as harmful is horizontal, [and takes place] between peers. When peers or friends disseminate information to each other, the privacy torts are much more difficult to apply."*

The appropriation, breach of confidentiality, publication of private facts and right to publicity tort might apply on a case-by-case basis, but proof of harm is difficult and incentives to sue are low. Experts do not agree on the applicability of the false light tort to privacy harms in Social Network Sites.

7.5. Facebook case study harms

In paragraph 4.5, I described how Social Network Sites try to make money from the identity-relevant information that users upload to their profiles. The SNS has incentives to create revenue for sustainable business model. In the Facebook case study, it became clear these interests are sometimes conflicting and lead to an unfair deal for users of SNS. The users have minimal bargaining power, because of network effects. If most of your friends are a member of Hi5, you cannot derive the benefits from their membership if you are a member of MySpace. You have to be a member of the Hi5 network. Recently, we have seen initiatives to open Social Network Sites to other websites, but most information (and therefore the most benefits) stays behind the walls of one website. Users cannot walk away from a deal that a SNS offers and there are no options for negotiation about which information is mandatory to provide. In addition, it is impossible to unbundle the services that SNS offer. As mentioned before, users can opt-out ex post of the pictures of them posted and tagged by others, but the harm is already done. Privacy settings give users some control over which information they disseminate to whom, but in general these privacy settings reflect the economic incentives of the SNS more than the preferences of the users.

The image that arises from the Facebook case study is that users have no or minimal control over the information they, or others, upload to the Social Network Sites. A good example of this is Facebook's Beacon, that collects information from partnering websites about users' behavior on these websites. Users can opt-out of the dissemination of this information to their friends, but they have no information on what Facebook does with this information, let alone have any control over these processes.

The Facebook case study did not offer any proof of dissemination of information to the government. As shown by the Beacon data analysis, it does prove that users are tracked online. Because the case study focused on the operations of SNS, it also did not provide any proof of dissemination of information to wrongdoers.

7.6. Risks identified by the IWGDPT

The International Working Group on Data Protection in Telecommunications, as discussed in paragraph 4.1, identified privacy risks for users of Social Network Sites which can be grouped into risks from insecurity, misleading practices by SNS and involvement of third parties (International Working Group on Data Protection in Telecommunications 2008).

The risks grouped as insecurity stem from the combination of hard- and software used on the Internet and Social Network Sites. Information is difficult to delete from the Internet, it does not have oblivion, the infrastructure of SNS is infamous for its data breaches and existing unsolved Internet and privacy problems as stalking and spam are magnified on SNS (International Working Group on Data Protection in Telecommunications 2008). The experts that I surveyed did not recognize these harms, probably because they focused on the new types of harms that originate from SNS. That is not to say that these risks should not be addressed, it means that they might be better solved on an Internet-wide base.

The IWGDPT (2008) clearly states that Social Network Sites engage in misleading activities because they have strong incentives to create revenue from the information that users upload. This creates risks for users such as online activity tracking and a lack of control over the usage of the data for secondary use, risks that were also identified by the experts. The financial incentives of a SNS are clear in the IWGDPT's risks (2008) that users eventually pay for the 'free' SNS with the sale of their data and the monetary incentives that SNS have to come up with new features that decrease users' control over their information.

The third group of risks originates from the involvement of third parties with Social Network Sites, as widely recognized by different experts. The IWGDPT (2008) mentions ID theft and misuse of the profile data by or sharing it via APIs with third parties, next to giving away more personal information than you think. The latter risk increases with new technologies such as Content Image Based Retrieval (CIBR), which matches identifying features of a location, such as a specific building, with information in a database (International Working Group on Data Protection in Telecommunications 2008).

Except for the privacy risks that are inherent to the usage of the Internet and deserve attention on a larger scale than just Social Network Sites, the risks that the IWGDPT identified have significant similarities with the privacy harms that the experts identified.

7.7. Risks identified by ENISA

The ENISA position paper on security issues in Social Network Sites (2007), mentions 15 risks for users of SNS. I will mention the ones here that are not mentioned by the experts or the IWGDPT or were not elaborated upon in the chapter 4 where the ENISA report was introduced.

As Gross and Acquisti (2005) mentioned, most profiles on SNS contain pictures of the profile owner that either implicitly or explicitly identify him. With the improved efficiency of face recognition algorithms and computing power (ENISA 2007), it will become fairly easy to connect a picture from an identified profile to a pseudo-anonymous profile. People may have very good reasons to create a pseudo-anonymous profile, for example because they do not want to be recognized on a dating site. Face recognition technology renders this impossible.

Not only can identification data be subtracted from the images that SNS users post on their profiles, Content-based Image Retrieval features on a picture, such as a background or a view from a window, can be used to determine the location where a picture is taken. This technology brings all the risks with it that unwanted disclosure of location data enables, such as real-world stalking, surveillance or blackmailing.

Account deletion is not always straightforwards in Social Network Site. Some websites, like Facebook, allow a user to deactivate his account, but deletion is a difficult and time-consuming process in which emails have to be sent to the Facebook support staff (Daily Kos 2007). Furthermore, information about a user might be disseminated through many profiles, because of tagging of pictures, discussion boards and other public spaces on other users' profiles. For a user, this leads to the loss of control over his own information, he can not present his moral identity as he wishes. ENISA (2007) argues that this might be in contravention with the European Data Protection Directive (Directive 95/46), because that Directive states that personal data should be deleted after it has been used for its purposes and that it should be kept in a form that identifies the data subject no longer than is necessary for the purpose.

ENISA also identifies security threats that are amplified by SNS, such as spam, cross site scripting, viruses and worms and aggregation of information in portal-like websites. In accordance with the IWGDPT, ENISA recognizes that spam can be disguised as normal SNS activities, such a promoting one's profile, writing on public spaces of other users and adding friends. Spam profiles and comments mostly provide links to spam websites, with for example pornographic content. Cross-site scripting has been a problem for a long period on websites that allow the posting of information to their website. Because this information can include links or code which refers to functions on other websites, malicious code that is run on other websites can be linked to a benign website that allows the uploading of such information. Social Network Sites aggregators provide a single point for a user to find all the profiles of his friends and to manage his own profile. This increases the risk of account breach, because with one central entry point hackers could gain access to multiple accounts (ENISA 2007). Furthermore, the aggregation of identity-relevant information on one place could lead to unwanted dissemination of information to different social spheres.

Phishing attacks involve the creation of a fake website or email, which resembles a real-world website or email (social engineering), to abuse the users trust in this website and inquire for logins and passwords. Spear phishing is a highly personalized form of phishing, directed at one person or a group of person, thereby increasing the trust. Social Network Sites increase the probability of and the damage done by phishing, as the relationships between users on SNS are (partly) based on trust (boyd 2005). By creating a fake profile of a trusted entity, phishing criminals can easily trick users into giving away vulnerable information, leading to information-based harm.

As I will elaborate upon in paragraph 8.2, Sophos created a fake profile for a SNS with the goal to connect to as many people as possible and obtain as much information from them as possible (Sophos 2007). The intrusion of Social Network Sites by malign actors could lead to unwanted information leakage. According to ENISA (2007), this forms the basis for other threats as disclosure of private information, phishing of information and spamming campaigns.

Finally, ENISA (2007) identifies the risk of corporate espionage via SNS. Because SNS show the connections between employees of a firm and allow employees to easily post information on their whereabouts to one central location, access to sensitive data and a complete list of stakeholders in the company are easy to acquire. This could lead to loss of sensitive corporate data and financial losses.

7.8. Conclusions

In this chapter, I have found that three groups of privacy threats for users of Social Network Sites. These groups are 'total information awareness', 'dissemination to wrongdoers' and 'no control over data'.

The first is information collection by tracking users' behavior on SNS and other websites. Experts believe that the government uses this information and this leads to information inequality and could chill free speech.

The second group is dissemination of information to wrongdoers, which results in damage from emotional distress (stalkers / predators / bullies), financial damage (identity thieves) and harm to a users reputation from the unwanted disclosure of embarrassing information.

The third group contains all of the harms relating to the lack of control that users have over their own information. In the Facebook case study, I also found that others can upload information about users without their consent. Users have an unfair deal with Facebook because bargaining is not possible because of locked-in network effects, users cannot opt-out of many Facebook services ex ante, but have to wait until one of these services is activated before they are notified that they can opt out.

Experts agree that American privacy tort laws are not much of a remedy against these threats. The appropriation tort could protect users against the unwanted use of their likeness; however, as long as courts struggle with the definition of damage and the monetary benefits remain low, its protection is minimal. The breach of confidentiality tort could play a bigger role in the future, as more and more services are added to Social Network Sites, some of them including confidential information (such as credit card information).

8. Examples of threats, incidents and damages

The eleven threats that experts identified in the survey materialize into incidents. These incidents, such as identity theft, cause damage to users of Social Network Sites. These damages vary significantly per incident. To gain more insight in how threats materialize into incidents and how these cause damage to users, I will deconstruct incidents from the eleven threats that experts identified with the framework of chapter five.

8.1. Total information awareness

Online activity tracking

Social Network Sites track the activities of their users on their own and other websites. An example of this is Facebook's Beacon, which is described in paragraph 3.5. Experts agreed that this online activity tracking happens on a large scale and could severely damage users. I will shortly discuss how this incident fits in the classification of privacy threats.

1. Facebook collected identity-relevant information from its Beacon partners, third-party websites such as eBay.com, which Facebook users did not expect, because those websites are in different social spheres. Users might buy products for their friends, and do not want those friends to know.

2. Initially, this information collection took place with no or at least a dubious form of informed consent, namely the 'toast pop-up', which made users angry because it appeared so short that users would not notice. The information from the video rental site Blockbuster, for example, could be embarrassing if disclosed.

3. A proprietary Facebook algorithm analyzed the information collected from the Beacon partner, and decided to which friends to disseminate the information based on your friends list. This algorithm decides what information is sent to your friends, and because users do cannot influence[46] this, it restricts their ability to write their own moral biography.

4. Because the data subject is unaware of how this algorithm works, a power relationship exists between Facebook and the user. The user might not want some information to be collected or disseminated, but has no influence on this. This is clearly against the OECD's individual participation principle (Gellman 2008).

5. The collected and processed information is disseminated to some people in the friends list and could cross different social spheres without the consent of the user. In this scenario, your beer drinking soccer friends could become aware of your academic life, or worse, the other way around.

6. This information could be embarrassing, and could harm the user once it is disseminated to wrongdoers.

7. Because the user has no influence on which information is being sent to whom, it restricts him in his moral autonomy. This differs from the harms under (3) and (4),

[46] However, a recent visit to Facebook shows that other users can choose to 'show more of this users' or 'show less of this user' in their own Newsfeeds.

because it does not prohibit users from crafting their moral identity, but from controlling the dissemination of their identity to others.

	Information-based harm	Informational inequality	Informational injustice	Moral identity / autonomy
Information collection	2. Information collection takes place without informed consent and this information could be embarrassing to users.		1. Information collected from third-party website, other social sphere.	
Information processing		4. Users are unaware of how this algorithm works.		3. Algorithm defines to whom the information will be disseminated.
Information dissemination	6. Information could be used to harm user, for example embarrassing information.		5. Information is being disseminated to friends, in various social spheres.	7. User not able to build his own moral biography.

Table 5: Framework applied to Beacon case

The infamous Beacon case shows clearly how collection of information by online activity tracking leads to damages for users of SNS. The lack of control of the collection, processing and dissemination amplifies these harms.

Government usage of information

Experts worry most about the government or law enforcement using the identity-relevant information on Social Network Sites to gather information about its citizens. However, it has always been hard to find proof and details of governments snooping on citizens. Only in 2006, did the New York Times discover that the US Government was tracking financial transactions of citizens and monitoring phone calls since September 2001 (Out-law.com 2006). However, NewScientist magazine found that research conducted by the University of Maryland and the University of Georgia was funded by the American Advanced Research Development Activity (ARDA), within the project 'An Ontological Approach to Financial Analysis & Monitoring' (Marks 2006 and Aleman-Meza et al. 2006). The goals of this research was to to connect offline information about scientific reviewing with online information regarding friendships and shared interests on SNS. The NewScientist found that the role of ARDA is to spend National Security Agency's budget on research that can “solve some of the most critical problems facing the US intelligence community” (Marks 2006). This program is a follow-up to the earlier mentioned Total Information Awareness program (Marks 2006). The link between the research and government interest is clear, but what does the research include?

The paper from Aleman-Meza et al. (2006) investigates a method to define and identify Conflicts of Interest (CoI) between authors of scientific papers, based on their co-authorship and information from a social network, described with the Friend-of-a-Friend (FOAF) markup language. A summary of their work is depicted in illustration 11.

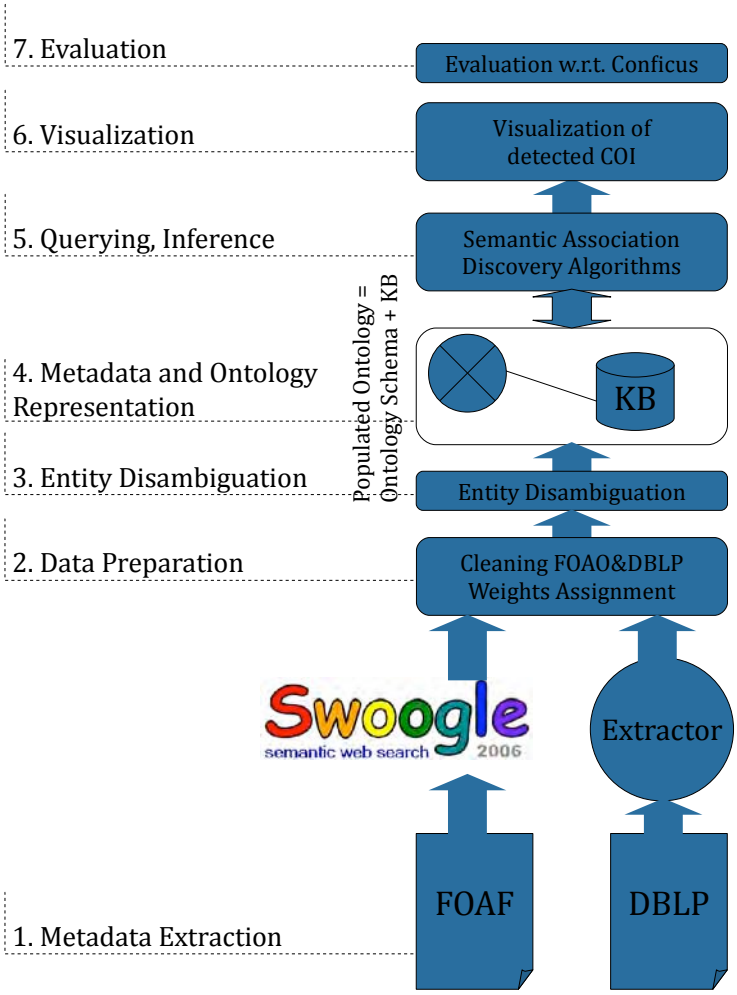


Illustration 11: Data collection and analysis from a SNS and another source. Adapted from Aleman-Meza et al. (2006).

Data collection from disparate sources and analysis of this data both separately and combined are important features of government use of information. The researchers from the University of Maryland and the University of Georgia mention that "privacy concerns prevent such [Social Network Sites] from openly sharing their data" (Aleman-Meza et al. 2006). The combination of information from disparate data sources with the information on Social Network Sites could lead to the identification of relationships between people based on financial transactions, sexual orientation, political preferences or any other type of information that is available on SNS.

They also mention that "[w]e acknowledge that there are privacy issues, [yet] a discussion on this is out of the scope of this paper." (Alemanan-Meza et al. 2006). However, the classification of harmful mechanisms from chapter 5 shows clearly where these privacy implications exist and what their impact is on users.

1. The government collects information from disparate sources. SNS are used to investigate specific properties that can only be found by mining SNS, such as sexual orientation, personal relationships or religious preference. This information is abundantly available on SNS. Information is collected from different social spheres, and the data subject may not want this information combined. The collection of this data from different sources also leads to what Solove names 'an architecture of vulnerability' (2004), because the government owns so much information about its citizens that they become vulnerable to various harms.

2. As depicted in step 4 in illustration 11, the information from the different spheres is aggregated in the Knowledge Base (KB), a central repository that combines the information into profiles of users. As Alemanan-Meza et al. (2006) mention, the aggregation of this data is not a simple process, as the data about individuals are of different quantities and qualities. Cases of 'false positives' are especially harmful with respect to the data subject's sense of informational inequality and the restriction of his moral autonomy. A false positive occurs when two data sets are identified as being from the same person, but in reality are not. Because it is difficult for the data subject to gain access to and correct this information[47], it could lead to serious problems with respect to informational inequality. The aggregated information of the data subject provides a basis to perform the Conflict of Interest analysis, as seen in step 5 and 6 in illustration 11. Based on this information and the analysis, a profile of the data subject with respect to his integrity and conflict of interest is established. This profile serves as a guideline for those that use it, and preempts whatever identity the data subject would present to others. This leads to a restriction of the data subject's moral autonomy.

3. Dissemination of this information could lead to more problematic situations for the data subject. This is not depicted in table 6, because it is in fact dissemination by a third party and not the Social Network Site. However, dissemination of this information to law enforcement agencies, for example, could lead to all four types of harms as defined by Van den Hoven (2007). Inaccuracies in the KB could lead to severe harms, especially when the information is disseminated to law enforcement. Recently, this has come to light in the case of Herring v. US (EPIC 2008). In this case, Bennie Dean Herring got arrested and searched by police officers based on a faulty record of him from another law enforcement agency. However, because the police now had a (faulty) reason to search him, they found other evidence that they used against Herring. Herring claimed that under the Fourth Amendment exclusionary rule this evidence should be suppressed (EPIC 2008). From a perspective of human rights, one could argue that arrests based on faulty information and the use of evidence gathered during illegal arrests could lead to an 'architecture of vulnerability' (Solove 2004), significant informational inequality (Van den Hoven 2007) between the data subject and the government, and can be used for repressing political views. The inaccuracy in government databases has also led to problems with voting for transgender people in the USA (National Center for Transgender Equality 2007), because a person's gender in the database does not match with their gender as visually observed, which could lead

[47] To combat this the OECD has devised the individual participation principle, which states that any individual should be able to request all the information a data collector has about him and should have means to correct any faulty data (Gellman 2008).

to a denial of voting rights.

	Information-based harm	Informational inequality	Informational injustice	Moral identity / autonomy
Information collection				
Information processing		2. The information from different sources and spheres is aggregated into a Knowledge Base (KB)		3. A representation of the data subjects CoI with other data subjects is created. This information can be disseminated to law enforcement.
Information dissemination		1. Information that is disseminated by disparate SNS and data sources is collected by the government		

Table 6: Framework applied to Conflict of Interest Detection case

Although Alemanan-Meza et al. (2006) do not mention any privacy aspects in their paper, the implications of government use of information from disparate sources are severe, as indicated by Herring v US (EPIC 2008). Other publications in this field, such as those from Hollywood (2004) and Sheth et al. (2005), suggest that these government tactics are not isolated incidents.

Recently, the Dutch police in Amsterdam released an oral press release (Novum/ANP 2008) stating that they started a pilot to look for criminals on Hyves. The Team Digital Expertise of the Amsterdam police force developed software that would identify suspects' online social networks and detect leads to other possible criminals. The software was tested on police officers, but if the pilot is a success, a roll-out on a national scale would be considered.

8.2. Dissemination to wrongdoers

Stalkers / predators / bullies

The Megan Meier case shocked the United States of America, and could have severe implications for privacy in SNS. Megan Meier, a 13-year-old girl from St. Louis, Missouri, committed suicide just after a 16-year-old, handsome MySpace friend told her "the world is a better place without you in it" (2008a). However, the 16-year-old boy was in fact a group of people who lived next door and had created a fake profile. Lori Drew, the mother of a classmate of Megan, participated in this hoax. She was indicted for violating the Computer Fraud and Abuse Act (CFAA) with one count of conspiracy and three counts of computer breach, namely creating fictitious profiles, sending abusive messages and soliciting personal information from minors (2008a). Experts agree that the damage from stalkers / predators and bullies on SNS can be very large.

Daniel Solove has a great collection of web logs about this case (2007d, 2007e, 2007f, 2007g, 2008c, 2008d), and specifically questions the decision of the prosecutors to charge Lori Drew with computer breach and violating MySpace's terms of use. If violating the 'terms of use' of a Social Network Site is considered a crime, then many people are criminals. Solove

(2008c) agrees with Orin Kerr that the CFAA was stretched too far by applying a SNS' terms of use to a criminal case. Specifically, he mentions that Drew's acts might be immoral, but not illegal (Solove 2008c).

Omissions in the law and the exact cause of the damage become much more clear when I analyze this case using the framework presented in paragraph 5.4. It must be stressed that this is not a legal analysis, but uses an ethical framework to identify the specific activities and harms. It is initially a descriptive framework, but it has normative implications when identifying the harms. In this case, three specific activities lead to harm, see table 7.

1. The disclosure of Megan's profile ID made it possible to contact her. Without this information, the group around Lori Drew could not contact Megan Meier and harass her. However, even though this information was used to harm the girl, it could also have been used in a benign way. It is therefore difficult to describe this as an information-based harm. Van den Hoven (2007) is quite clear about this: if identity-relevant information is insufficiently protected, it could harm people. According to him, "[I]n information societies, identity relevant information resembles guns and ammunition." I argue that because the contact via Megan's profile took place on a very intimate and emotional level, it could have led to severe harms.

2. The information of Megan's profile moved from a youth sphere to an adult sphere, which contained Lori Drew. The involvement of Drew in the harassment is what most people find appalling (2008a). Megan's contact information was not used by a peer, but by an adult with malign purposes, who, considering her age, could have harmed Megan more easily. Solove asks in one of his posts (2007d): would people feel differently about this incident if it were teenagers who harassed Megan? I argue they would, because the transfer of information to the adult sphere and the involvement of a more mature person in the harassment especially appeals to our feelings of morality.

3. Finally, the woman and girl posted harmful remarks towards Megan on her MySpace profile. The information was uploaded by the group and collected by MySpace. This information, and especially the final remark, was posted with the intent to harm Megan. Solove (2008d) states that it is hard to prove that these remarks led directly to the suicide. However, it is very clear that the remarks were made to harm Megan, and therefore part of information-based harm.

	Information-based harm	Informational inequality	Informational injustice	Moral identity / autonomy
Information collection	3. Harmful remarks towards Megan are uploaded to MySpace (collected).			
Information processing				
Information dissemination	1. Disclosure of Megan's profile ID makes contacting her possible.		2. An adult, from a different social sphere, contacts Megan.	

Table 7: Framework applied to Megan Meier case

With this analysis, I have showed that the specific harms in the Megan Meier case derive from the possibilities that identity-relevant information is used for harm, and the movement of this information through different spheres.

Identity theft

In a research project on identity theft, security software manufacturer Sophos created a fake Facebook profile, Freddi Staur, and asked 200 Facebook members to become his friend (Sophos 2007). 41% of the users that they contacted gave away identity-relevant information such as email address, date of birth and phone number to a green frog called Freddi Staur, who divulged minimal information about himself (Sophos 2007). Graham Cluley, senior technology consultant at Sophos, says "while accepting friend requests is unlikely to result directly in theft, it is an enabler, giving cybercriminals many of the building blocks they need to spoof identities, to gain access to online user accounts, or potentially, to infiltrate their employers' computer networks." (Sophos 2007). Experts agree that identity theft is one of the most harmful activities a SNS user could encounter.

That the information on SNS is easily accessible and can be used by stalkers is something Samer Elatrash experienced in real life (Elatrash 2007). He is, within certain groups, a well-known pro-Palestinian Israeli who signed up for Facebook because his friends asked him to. After two weeks, he found out that someone copied information from his profile and started a fake profile under his name. The impostor was a member of various dubious Facebook groups and clearly tried to ridicule and harass Elatrash.

Applying the framework to his case, I can identify the following harmful activities, which are also depicted in table 8:

1. Someone was able to get access to Elatrash's real profile, and copy the disclosed identity-relevant information to create a fake profile, very similar to his profil. Real friends told him they thought he had two profiles, and could not tell the fake from the real one (Elatrash 2007). With the information from his real profile, the impostor was able to create a harmful fake profile, thus this is information-based harm.

2. A few of Elatrash's friends, and a few dozen other members, thought that the profile really represented him (Elatrash 2007). The impostor was even able to shop for

apartments in his name. This certainly restricted Elatrash in his moral autonomy, because he was not fully able to shape his own online identity anymore. As becomes clear from this case, information about other people is always 'knowledge by description' (Van den Hoven 2007) and Elatrash's friends therefore had difficulties with distinguishing between the fake and real profile.

3. The difficulties in distinguishing between fake and real made the untrue and embarrassing information posted on his profile even more harmful. It started with mentioning the Quran as his favorite book and the “sleazy pop star” George Wassouf as his favorite musical preference (Elatrash 2007). But soon, the information became an exaggeration of his political viewpoints, when the impostor joined a group that sought the enforcement of Islamic law and the restoration of the Caliphate. When the impostor used the profile to look for an apartment, this information was disclosed to even more people. The fake information was deliberately posted to harm Elatrash. After he notified Facebook and identified himself as the real Samer Elatrash, Facebook deleted the profile.

	Information-based harm	**Informational inequality**	**Informational injustice**	**Moral identity / autonomy**
Information collection	3. Harming Elatrash by disclosing untrue and damaging information.			2. Creating a profile similar to that of Elatrash
Information processing				
Information dissemination	1. Disclosure of Elatrash's profile information.			

Table 8: Framework applied to Samer Elatrash case

Elatrash became the victim of a very special case of identity theft, one in which the identity thief tried to impersonate him by means of a fake profile and harm him by disclosing untrue information about him. ENISA (2007) mentions that this form of identity theft, also known as profile squatting, is occurring more frequently. Social Network Sites enable profile squatting because the abuse can be better targeted at the people who are most likely to notice it, SNS users assume that a profile is created by the person who it purports to represent, the target of the attack might not be able to access the profile and SNS perform only weak registration of new users (ENISA 2007). With help of the framework, I have shown that Elatrash's moral autonomy was damaged and the impostor harmed him by both collecting his real information and disseminating fake information.

Damage to reputation because of disclosure

In one of my blog posts (2008i), I talked about the case of Jodie and Amanda Hudson. In May 2008, British newspapers and bloggers mentioned that a party in Spain was thrashed because the organizer, Jodie Hudson, posted the details of the party on her Bebo, Facebook and MySpace profiles. However, the information on the profiles turned out to be untrue and an infuriated mother, Amanda Hudson, sued newspaper the Independent for defamation and breach of privacy (Verkalk 2008). Obviously, the Hudsons were damaged by the disclosure of

the untrue information about them, but the untrue information was posted by Jodie Hudson herself. Experts rated this form of 'damage to reputation because of the disclosure of information via SNS' as having a very negative impact on users.

This case is being followed by legal and privacy experts worldwide, as lawyers say the outcomes of the court case "could place a duty on all second-hand [information] users to establish the truth of everything they want to republish from such sites." (Verkalk 2008). Again, the framework makes clearer where the exact harms are initiated.

1. The information that Jodie Hudson posted on her website seems not to be intended for newspaper publication. As Matyszczyk (2008) wrote, why did nobody ask Jodie why she did it? Does not everybody know that everything that you post on a SNS can be read by every person on earth? It is probable that, as with the unwanted dissemination case at Oxford that I will mention in paragraph 8.3, Jodie Hudson did know that all her Bebo friends could see the post, but did not expect the information to transverse into another social sphere, that of the institutionalized media. And it is exactly this dissemination to spheres beyond Jodie Hudson's control that led to the feeling of breach of privacy. Refer to the discussion of Strahilevitz' Social Network Theory (2004) in paragraph 4.2 on why this expectation of privacy is reasonable.

2. Once the newspapers gained access to the information, they published it without checking the facts. Jackson (2008) mentions that Sky News, UK Times Online, Daily Mail Online and The Register reported on the story. It would be easy for the reporters to check the truthfulness of the information, as it was reported that the police was involved. Reporters should have been able to check the information against police reports. But this carelessness led to the dissemination of untrue information about the party to millions of people. The Hudsons were unable to restore this image, thereby affecting their moral autonomy. However, a ruling against the newspapers would pose extra efforts on media to check for the truthfulness of facts. Although such a ruling would have been highly controversial, checking the facts would have been relatively simple in this case.

	Information-based harm	**Informational inequality**	**Informational injustice**	**Moral identity / autonomy**
Information collection				
Information processing				
Information dissemination			1. Untrue information, intended for people on the friends list, travels to the media sphere.	2. Newspapers disseminate information on large scale, inhibiting moral autonomy of the Hudsons.

Table 9: Framework applied to Jodie and Amanda Hudson case

The Hudson case is complicated because it involves self-disclosure of untrue acts, which is amplified by the dissemination through a Social Network Sites and institutional media. However, the travel of information through different spheres and the inability to shape the

true picture of the party, are concrete harms made visible by the framework.

8.3. No control over information

No information over / control over secondary use

In paragraph 5.3 I discussed various viewpoints regarding the Social Ads in Facebook's Beacon. This part of the Beacon program shows your friends what you have bought on other websites and uses your likeness for viral marketing. I also described the reactions of McGeveran (2007 and 2007a) and Solove (2007b and 2007c), who see this as a clear-cut case of appropriation, or using someone else's likeness for your own benefit. In essence, the appropriation tort protects people against the secondary use of their likeness or buying behavior.

Although the discussion of Solove (2007b and 2007c) and McGeveran (2007 and 2007a) focuses on the question whether or not Social Advertising is a form of appropriation, their analysis shows clearly how users are harmed. McGeveran (2007) mentions that users are only asked in general if they want to share information, not if they want to feature their name and picture in an advertisement. This clearly violates the OECD's purpose specification (Gellman 2008), because users are not informed specifically where the collected information will be used for. Furthermore, the users have no influence on when and to whom the information is disseminated. Thus this restricts them from building their own moral identity, as an advertisement for any embarrassing product seen by their friends could harm them.

1. The user has not control over which purchasing information is disseminated to whom, thereby restricting his moral autonomy and prohibiting him from creating his own moral identity.

2. The algorithm that Facebook uses to determine to whom the information is disseminated is proprietary, so the user has no information or control regarding to whom the information is disclosed. The purchase of products that are considered taboo could well end up with one's in-laws, who might have been added to the friends list last summer in order to see vacation photos.

	Information-based harm	**Informational inequality**	**Informational injustice**	**Moral identity / autonomy**
Information collection				
Information processing				
Information dissemination			2. Because the user has no control over the destinations, information could end up in social spheres were the user does not consent to.	1. When a user buys something from a Beacon partner, this is disseminated to (a part of) his friends list, alike an advertisement.

Table 10: Framework applied to Facebook Social Ads case

The Facebook Social Ads case clearly shows problems with the interpretation of

informed consent, and the (economic) incentives SNS have to define it as broadly as possible. Furthermore, it shows that the business model of the Social Network Site makes it necessary to keep certain parts of their operations confidential, such as the algorithm that defines to whom the information will be disseminated. However, this information is essential for users' moral autonomy and to prevent informational injustice. This is especially concerning, because experts agree that this secondary use without consent is occurring on a large scale.

Sharing / selling of data to third parties

The Social Network Sites Facebook, MySpace and Hyves offer users the possibility to add applications from other software developers, third-party applications, to their profile. These applications compare your birthday with those of your friends and compare horoscopes, for example. When a Facebook user adds such an application to his profile, he needs to check a box saying "allow this application to [...] know who I am and access my information." (Soghoian 2008b). A reasonable user would expect that the birthday application gets only access to the birthdays of himself and his friends (Soghoain 2008a). However, Adrienne Felt from the University of Virginia found that 91% of the third-party applications get access to far more information than they need to perform their function (Felt 2008b). According to experts, the sharing of information between SNS and third parties is happening on a large scale.

In interviews with Christopher Soghoian and Adrienne Felt, I asked them more about the implications of this over-sharing of information with third parties. Soghoian (2008a) explained that in computer science, applications and functions are given as little permission and information as needed to perform their function, known as the 'least authority' principle. With the checkbox that Facebook offers now, users are left with only one choice: accept the sharing of all their information or not. He also mentioned that is very hard to identify any harm from this, as it is unclear what information the third-party developers store and what else they do with the information. From a perspective of information-based harm, most of this information can be used to harm users. Once this information has left Facebook's servers, the SNS has no way to protect this data against malicious intent (Soghoain 2008a). An even more troublesome development is that the third-party developers also get access to the information of your friends, even if they have set their profile to private (Soghoian 2008b). Chris Soghoian (2008b) says:

> *"A lot of Facebook users [..] put their profiles on private. So public users can't see the profiles. This is a concrete step that many users take to protect their privacy and [a conscious step] against stalkers. If you add a third-party application, the developer will be in your shoes and see whatever information you can see, including the private profiles of your friends. That is very troubling, because your friends didn't consent to this and don't know this could happen."*

When Hyves opened up its network for third-party applications, it at least showed exactly which information the third-party applications could access in the notification box. However, as I mentioned in an April interview (TROS Radio Online 2008), third-party application developers are still able to get access to the tips, blogs and status messages of your friends, without their consent.

It is worthwhile to mention here that Adrienne Felt (2008a and 2008b) has designed a

solution which will at least not send referential information about users to third parties. In an interview (Felt 2008b), she explained that the third parties explicitly have to request the fields of information they want to receive. Upon this request, Facebook disseminates the information to the third party application providers. Facebook could easily operate filters for the requested information, although Adrienne Felt admitted that she believes that "a finer grained access control list [for users] won't solve the problem" (Felt 2008a) because users do not understand in what ways their information is shared, and would therefore never be able to make a fully informed decision (Felt 2008b). The solution that she promotes is called privacy-by-proxy, and comes down to replacing the unique (referential) identifier of the user with a randomly generated key. Facebook sends this randomly generated key to the third party developers instead of the referential identifier, which restricts them from knowing the real identity of the user. When Facebook gets the data that is processed by the third party back, it can substitute the randomly generated key with the unique identifier and disseminate the information to the user (Felt 2007a). Although I have concerns about privacy-by-proxy, specifically the re-identification by aggregating attributive data of the user, this solution at least shows that the same level of security as via the web interface can be implemented for the third party applications (Felt 2008a). Gross and Acquisti (2005) mention that re-identification, especially by comparing photos with personal information with photos without this information, is one of the privacy risks for users of SNS.

Schematically, it looks as follows:

1. By adding a third-party application to their profile, users consent to giving third-party developers access to all their information and all the information they can see from their friends (Soghoian 2008a). This information also includes what Felt (2008b) calls private information, and what I see as information that could harm a user severely (sexual orientation, political preferences, birth date, state of birth).

2. Third-party applications have access to much more information than needed, which is opposite from what users expect (Felt 2008b) (see also the interview with John Verdi, 2008). Information travels outside the expected social spheres, such as to third-party developers who gain access to identity-relevant information that your friends only intended for you to see.

	Information-based harm	**Informational inequality**	**Informational injustice**	**Moral identity / autonomy**
Information collection				
Information processing				
Information dissemination	1. More information is disclosed than needed to third parties, even sensitive information.		2. Users cannot expect that a birthday application gets access to photos, because this information resides in different social spheres.	

Table 11: Framework applied to Facebook Third-party Applications case

The Facebook Third-Party Applications case shows that users' expectations of privacy and expectations of an application's function on Social Network Sites is very different from what happens in reality. Also, it shows that SNS like Facebook do not have the same standard of security on every part or function of their website.

Unwanted dissemination of information to others / groups

Oxford University in England has a disciplinary body, the proctors, which has admitted to using Facebook in order to find evidence of students breaching the University's code of conduct (Gosden 2007). A common post-exam tradition at Oxford is 'thrashing', or covering your friends with flour, confetti or even raw meat or octopus. The University does not approve of this and students have been receiving fines of 100 pounds and a prohibition from graduating (Knight 2007).

Recently, students that have posted pictures of these activities on their Facebook profiles have been disciplined. Of course the actions of these students are against the code of conduct of the University, but the question here is why the students got so upset when the University used this information to discipline them. As one of the students, Alex Hill, says: "I don't know how this happened, especially as my privacy settings were such that only my friends and students in my networks could view my photos. It's quite unbelievable and I am very pissed off, [I] just hope that no-one else gets 'caught' in this way." Part of the amazement comes from the fact that the students did not expect that the University would be able to access the information on the SNS. Experts agree that the unwanted dissemination of information to others is something that happens on a large scale.

1. The students posted information online, only for their friends and fellow students to see. Because Facebook started as a SNS for college students only, it created the image of being only accessible to students from the same college. The last thing that students expected was that the information that they posted online could be accessed by the University's proctors. The pictures of 'thrashed' students would not have the function of amusement, but of evidence for the proctors. The function of the same pictures is different in one social sphere from another and this harms the students.

2. Oxford University gets access to more information about students than they desire, and this creates what Solove (2004) calls an 'architecture of vulnerability'. Although Solove applies this concept to the government, the prohibition of graduating (Gosden 2007) has devastating effects on students. The outcomes are the same: the students have to watch what they are saying, doing and posting online, because they do not know what information the university can access and use to enforce its code of conduct. This could at least chill free speech, for example, as criticizing the policies of the university on Facebook will be less appealing.

	Information-based harm	Informational inequality	Informational injustice	Moral identity / autonomy
Information collection				
Information processing				
Information dissemination		2. Architecture of vulnerability, because University has much information about student.	1. Dissemination of information intended for friends to sphere of University proctors.	

Table 12: Framework applied to Alex Hill case

The Oxford / Alex Hill case demonstrates what happens if information is disseminated to groups that we do not want to disclose the information to. This could surprise and severely harm us, especially if we do not expect the information to end up in another social sphere with a different function.

Posting of information by others

In August 2008, a court case in The Netherlands quickly gained attention from the media. Marie-Jose Klaver from NRC Handelsblad (Riphagen 2008h) wrote about a Dutch judge in Drenthe who convicted a woman for libel because she wrote that her ex-husband was a pedophile on her private profile on Hyves (ANP 2008). Earlier, a judge in Zwolle decided that posting negative untrue information on Social Network Sites constituted libel (Doornbos 2008).

The decision by the judge -- classifying the private profile of the woman as public -- is controversial, but there are also other ways to analyze the case. By posting the untrue information on her website, the woman intentionally wanted to harm the man by spreading information through different social spheres and making it harder for him to construct his own moral identity. Structuring this according to the framework looks like in table 13.

1. The woman intentionally posts information about her ex-husband that is untrue and would put him in a false light on her private profile of SNS Hyves. With this disclosure, she tries to harm her ex-husband.

2. Because people on her friends list have access to her profile, the information becomes disclosed to different social spheres, possibly to ones that the data subject does not belong to. Because she posted the information without his knowledge, he was not able to consent to this.

3. The data subject has limited means to constrict the posting of this information and the dissemination of this information through different social spheres. The data subject claims that his neighbors were acting strangely towards him, as they supposedly have read the post on the website (ANP 2008).

	Information-based harm	Informational inequality	Informational injustice	Moral identity / autonomy
Information collection	1. Intentionally posting of untrue information about another on a private profile.		2. People in friends list, unrelated to data subject, have access to the information.	3. Because the data subject has limited options to correct information, he is restricted in his moral autonomy.
Information processing				
Information dissemination				

Table 13: Framework applied to Dutch 'pedophile' case

In this case, the judge decided on libel because he deemed the private profile of the woman as public. However, there are plenty of immoral activities that make up this case. Such as the posting of identity-relevant information for disclosure to different social spheres without the consent of the data subject. All of this with the goal to create an untrue image of the data subject.

The posting of information by others is very common on Social Network Sites, as Facebook's tagged function or Hyves' 'gespot' functions show. ENISA (2007) mentions that "privacy may be under even greater threat from images posted by other". In their report, the researchers also mention that the metadata that digital cameras embed in their photos, linked to the address data in the warranty cards, can lead to identification and threats a user's privacy. Experts agree that this happens on a large scale and could severely harm users of SNS.

8.4. Others

Displeasure from being monitored

Users of Social Network Sites usually express concern after new features are implemented that do not comply with their expectations. The dissatisfaction stems from the users not being consulted or informed about the new feature and the lack of control they have over their own information. On the other hand, the users do want to disclose their identity-relevant information online, which leads to what Barnes (2006) calls a 'privacy paradox'. When Hyves introduced some new features which were very similar to Facebook's Newsfeed and Facebook's Third-Party Application Integration, users created a website 'against the change' (2008c).

The focus of annoyance was that Hyves controlled who gained access to your profile, that the Newsfeed (called Buzz) would show others what, where and when you were doing something and that kids would have difficulties with understanding the implications of this feature. When the protestors were interviewed on Dutch National Radio (TROS Radio Online 2008), they mentioned that they were annoyed by the fact that they were monitored by Hyves without having any control about this.

I discussed on the radio (TROS Radio Online 2008) that the public / private feature led to much displeasure because users felt watched or monitored by Hyves. Additionally, if users put

their profiles on private (so users from outside their friends list could not see their profile), they would also not be able to view profiles of people outside their friends list (2008d). Although some users appreciated the reciprocity of the solution, most disagreed with the lack of control and the monitoring by Hyves.

1. When users decide to put their profile on private, they can only see other profiles from their friends list. Hyves controls the dissemination of information from other profiles, and when your profile is set to private, you cannot see those profiles.

2. This results in an unfair situation for the user, as Facebook has more information about which users have put their profile on private and about the friends of your friends network. If you put your profile to visible to friends-of-your-friends, you would be able to see those profiles. As a user has no idea how many friends-of-your-friends could then be able view his profile, it is difficult for users to make an informed decision.

	Information-based harm	**Informational inequality**	**Informational injustice**	**Moral identity / autonomy**
Information collection				
Information processing				
Information dissemination		2. Hyves has more information about whose profile is private, and can thus force users to make their profile public.		1. Users chose to not disseminate information to others, but are now restricted in the dissemination of information from other private profiles.

Table 14: Framework applied to Hyves case

The Hyves case is an interesting one, because it is the only case in which a Social Network Site restricts the dissemination of identity-relevant information to other users. However, because the SNS has more information about which users have set their profile to private and users have no bargaining power, the users who want to see other people's profiles without adding them are left no choice but to set their profile to public. This leaves the user with a feeling of helplessness and being monitored, something that experts recognized is happening on a large scale.

Aggregation of data / profiles of users

Social Network Sites consist of user profiles that form the central places where users' information is aggregated. Users choose to display their personal preferences, pictures and contact information in one central place. boyd and Heer (2006) mention that profiles have become a "common mechanism for presenting one's identity online", but also include explicitly social information such as links to friends and testimonials which help shape the representation of others on a SNS. This information is very valuable to marketers, as it contains all the ingredients for viral marketing, as described in the Facebook Social Ads case.

In 2006, Chevy Chase signed a one-year contract with Facebook that made it the exclusive credit-card sponsor of the SNS (Simon 2006). This contract included the creation of a special Chevy Chase network, which interested users could join, and a reward program for members of this group. Furthermore, undergraduate students could participate in an ambassador program that claimed to "help shape this new credit card program by providing us with your feedback, gathering information and representing the wants and needs of your peers" (Simon 2006). Clearly, the Chevy Chase network, the rewards program and the ambassador function served the goal of getting more information about users' preferences.

Van den Hoven (2007) mentions that the inequality of information between a consumer and a producer could lead to price discrimination. Earlier, I raised the problem of whether upward price discrimination is legal. Chester (2007) provides an extensive survey on profiling and the marketing industry in the United States. He mentions that SNS like MySpace and Friendster use the advertising technology from Doubleclick. His investigation into Doubleclick makes a good example of how targeted advertising works. Social Network Sites also make use of the services of advertising networks to show advertisements to their users.

1. Information about your surfing behavior is collected from different websites by using cookies or web beacons. Cookies are little bits of information that a web server places on your computer and can be stored indefinitely. For example, a cookie would be placed if you were searching for second-hand computers on MySpace, and would be updated with every new search you do on MySpace's marketplace. Web beacons are small pictures (commonly in GIF format) that have the same function, but could work across multiple websites. These were also used in Facebook's Beacon feature.

2. The information that is collected from multiple sites (in Doubeclick's case information was also collected from the SNS, MySpace) is aggregated into a profile in a database. This profile or 'digital person' (Solove 2004) represent the preferences of your real-life character and serves as a basis for ad-serving. Doubleclick calls this the Dynamic Advertising and Targeting Technology (DART) (Chester 2007).

3. Based on the collected information, marketers will decide what ads you will receive and might consider you for an 'upsell', and try to get you to spend more (Chester 2007). This is possible because the marketing company has more information about your behavior and that of people with similar preferences than you have.

4. The information about what ads to serve you next is disseminated to an ad-serving company, or in the case of DoubleClick, it presents ads itself on the networks of MySpace and other partners (Chester 2007).

	Information-based harm	Informational inequality	Informational injustice	Moral identity / autonomy
Information collection			1. Cookies or web beacons collect information from different sites / spheres.	
Information processing		3. Marketers know what your interests are and might pick your for an 'upsell'		2. Information stored in database for profiling and ad-serving.
Information dissemination		4. Because your behavior can be predicted, you receive targeted ads.		

Table 15: Framework applied to Doubleclick / MySpace case

The DoubleClick / MySpace case shows how collecting information from different spheres, and combining it with information from Social Network Sites into a profile of a user, can lead to targeted advertising, which has as goal to have you spend money. This happens on a large scale, according to experts.

Recently, an application called 'We feel fine' has appeared on the Internet that collects expressions of feelings, including text and images, from different parts of the web and combines this into a database. The application then collects identity-relevant information from your blog, for example, and ties this information (location, name, age, gender) to the feeling (We feel fine n.d.). As an extension on this work, Nitesh Dhanjani, senior manager and leader of application security services at Ernst & Young, and Akshay Aggarwal, Microsoft InfoSec team's practice manager for North America, are developing a "proof-of-concept" tool that analyzes a feed from peoples' various online presences" (Mills 2008). This severely intrudes upon a users' moral autonomy and shows the informational inequality between the marketing industry and the users of SNS.

8.5. Conclusions

In the previous chapters, I have analyzed Social Network Sites, their behavior and their interests. I have developed a taxonomy to address privacy harms in SNS through the use of identity-relevant information, and experts have identified and expressed their concerns about specific activities. In this chapter, I have shown that all this work is built upon real issues, and that the taxonomy proves very useful for analyzing the real-world examples of privacy harms for users of Social Network Sites. When I started writing this chapter, I hypothesized that it would be difficult to come up with privacy threats that materialized into incidents, but the contrary is true.

9. Reflections

9.1. Discussion on methodology

Research on privacy issues within Social Network Sites has not advanced to the stage where there are broadly recognized frameworks to identify privacy threats. Nor are there methods that guide policy makers by deconstructing privacy incidents and finding exactly where harm takes place. Thus, building on the relevant work of recognized privacy experts, I developed my own framework to analyze the privacy incidents in Social Network Sites. This way a framework was developed that is customized to investigate privacy issues in SNS. However, this also has its disadvantages: the framework could suffer from bias towards advocating privacy protection, the validity of the framework needs to be assessed and the external validity of the framework is uncertain.

Because the framework is designed to identify those activities that harm the privacy of users and to find the reasons why this harms users, it surpasses the point of discussion about whether these privacy incidents are actually occurring and whether these are harmful. One of the respondents of the survey noted that he believed that the increased accessibility of identity-relevant information in Social Network Sites did not posed any privacy risks, as users decided themselves to disclose the information. Although I recognize the problem with self-disclosure and the inability of the law to protect users against this, the 11 examples of privacy harms in chapter eight speak for themselves. The balance between analysis and advocacy is delicate, but there is sufficient proof that at least a group of users of SNS is unhappy with the privacy risks they are subjected to. The lack of information on how SNS internally operate increases the difficulties for an objective analysis. However, the application of moral reasons to restrict access to identity-relevant information balances this out. Because these reasons have a universal applicability, the balance between advocacy and analysis is well-struck.

It is difficult to validate a framework that is constructed for such a specific purpose. The application of the framework to the eleven privacy incidents (in chapter 8) that Internet and privacy experts identified, provides eleven good test cases for the framework. In all those cases the framework gave meaningful explanations of why damage was incurred and which activity triggered this. The external validity of the framework might be more difficult to assess, because it was constructed for such a specific purposes. However, the moral reasons to restrict access to identity-relevant information have been applied to the Dutch Electronic Health Record (Manders-Huits and Van den Hoven n.d.) with success. Furthermore, the activities that data holders perform with identity-relevant information are derived from American laws and various cases by Solove (2006). There was no time in this study to apply the framework to other areas where informational privacy is deemed important. This application is feasible, and it is recommended to apply the framework to another area of Web 2.0 Internet services.

Finally, the ratings of privacy harms on ‘probability of occurrence’ and ‘negative impact on users’ should be seen in relative light. It has proven to be very difficult to get a large response from privacy and Internet experts and quantify their ratings. This means that the ratings should be seen in a different light, which still renders them useful for policy makers. With respect to the most important areas for privacy threats, the three focal areas can be easily identified and the measurements do allow for a ranking of the most harmful threats. This provides policy makers with a useful prioritization.

9.2. Discussing the substance

Research agencies and scholars as ENISA, the IWGDPT and CIPPIC have written on privacy threats for users of Social Network Sites. These issues play an important role in Daniel Solove's 'The Future of Reputation' (2007). This thesis aimed to construct a framework that could be used repetitively for analyzing privacy risks for users of SNS and identifying preventive measures. It differs from the work of others, because it does not provide off-the-shelf solutions, but a framework that provides customized solutions for every incident. The framework is just a classification in three activities and four moral reasons to restrict the activities and therefore a model of reality. The three activities can be clearly distinguished in SNS, but the four moral reasons might overlap and are not mutually exclusive. I mentioned before, in paragraph 2.4, that the freedom of moral autonomy is always at stake whenever a digital representation of a person is employed. In an interview with Van den Hoven (2008) I asked him what he thinks about the loss of control over information that a digital representation results in. His answer was that[48]:

> *"The loss of control [over your own identity-relevant information] is certainly important. However, the question is to which extent you can prevent this. People are continuously not aware of information [about them] that could harm them. [..] But if a relative has information about me that is relevant for me, he should disclose it to me. Then it becomes morally relevant if this information is disclosed to me [and within my control]. "*

Sometimes the moral reasons are not easy to distinguish and even intertwined. Does the collection of information from secondary sources mean that users are harmed by informational injustice? Or does it mean that the SNS has the power to collect information about users however it wants, thereby causing harm because of informational inequality? The answer to this question is not easy, and should be taken into account when validating the framework externally. Applying the framework on a case-by-case basis is a process that should be carefully executed. This makes it more difficult to use the framework, but it is not difficult to apply and there is no reason to believe that it would not be workable for policy makers.

9.3. Discussion of the policy process

The outcomes of this thesis form the input for a policy development process that is aimed at minimizing the privacy risks for users of Social Network Sites. Before going into how the policy process should be developed, it is worthwhile to create an analogy with the guidelines for data protection in the direct marketing and financial sectors. In the Netherlands, both sectors have guidelines that restrict their activities with respect to information collection, processing and dissemination. A similar set of guidelines could well be used to regulate activities that threaten the privacy of Social Network Site users. Although both guidelines follow the OECD's Fair Information Practices (see paragraph 4.2) closely, they also distinguish sensitive data, as derived from the European Union's Data Protection Directive. Much of this sensitive data, such as religious, political and sexual preferences, are abundant on Social Network Sites. The Dutch guideline for Direct Marketeers (Federation of European Direct Marketing n.d.) mentions that for the collection of sensitive data explicit written consent from the data subject is required. However, "[when] the data have been manifestly made public by the Data Subject" (Federation of European Direct Marketing n.d.),

[48] Translated from Dutch by David Riphagen.

this is not required. The publication of identity-relevant information on SNS by users has led to problems specifically when the information travels to different social spheres. Because SNS promote the uploading of identity-relevant information and the dissemination of information through different social spheres, it deserves recommendation to assess how the explicit written consent for sensitive data could translate to Social Network Sites.

The analogy between banks, insurance companies and Social Network Sites could lead to good insights for policy measures. Both organizations collect information that is not primarily needed to support their business processes. Banks might get access to information about labour union membership, if periodic payments are processed via the bank. Social Network Sites obtain much more information about user's preferences than is needed to provide the service of social networking. The Dutch guideline for data protection in financial institutions (Gedragscode Verwerking Persoonsgegevens Financiële Instellingen) mentions that any information that is needed to execute financial transactions between organizations or people, but not primarily needed by the bank to support its processes, must be considered confidential. Banks cannot use this information for marketing purposes. A similar distinction between information that is primarily needed and information that is not primarily needed for business purposes provides a good starting point for designing policy measures for the protection of privacy in Social Network Sites. Although the creation of guidelines for information processing in SNS is feasible from a government perspective, the free flow of information creates revenue for SNS and their incentives are opposite the regulation of these information flows. Therefore, voluntary participation of Social Network Sites in this process does not seem feasible. Governments could however use the threat of unilateral regulation from their side if SNS do not participate in the policy development process.

The rational model of the policy process, as described in paragraph 1.3 and illustrated in illustration 1, consists of the following five phases: (1) Collection of objective information; (2) Information analysis and normative evaluation; (3) Policy development; (4) Policy implementation and (5) Policy evaluation. The outcomes of the information analysis and normative evaluation are described in the previous paragraph. The most important privacy threats for users of SNS are 'Total Information Awareness', 'Dissemination of information to wrongdoers' and 'No control over your identity-relevant information'. How can policy makers design preventive measures to preempt these threats and what should these measures look like?

The prioritization and grouping of privacy harms provides an agenda of subjects for policy makers to address. Initially, policy makers should focus on creating a better and more fair balance regarding what the information SNS can collect from which sources. The guideline here is always that a user should be well informed and fully consent before collection takes place. The Fair Information Practices of the OECD (Gellman 2008) can be used as a guidance. Furthermore, there should be a discussion on government access to the collected information, and especially on how government abuse can be prevented. The framework to deconstruct privacy incidents can be used to identify how exactly preventive measures should minimize the threats for privacy.

The preventive measures should influence the pay-offs in the matrix in table 1 in such a way that the outcomes become desirable. Social Network Sites derive monetary benefits from sharing users' information with advertisers for profiling. As long as these benefits are greater

than the costs SNS incur from regulation, and as long as these benefits can only be obtained by sharing users' information with advertisers, SNS have no incentive to stop these harmful activities. There are two options to influence this trade-off:

1. Impose regulations on sharing information of users with third parties, such as advertisers, and impose fines on these activities. The fines should be of such an amount that the costs for SNS of sharing become higher than the benefits. Also, regulations could impose more privacy-friendly default settings for SNS and give users more control over their information, as proposed by Solove (2007a). Governments have the power to institute such regulations, but it is doubted whether world-wide regulation can be easily achieved. It it also difficult to monitor and enforce compliance of Social Network Sites with such regulations.

2. Proposals to create revenues for SNS that are less invasive on their users' privacy. For example, value-added services that do not use identity-relevant information without the explicit consent of the user could be promoted. It should also be considered if and how much money users are willing to pay for a SNS that does not share their information without their consent. It is questionable if a new equilibrium in the sharing of identity-relevant information and coupled revenues can be achieved, as nothing is known about users' willingness to pay for better privacy protection. Proponents of market solutions will argue that if users were willing to pay for better privacy protection, this would already be offered by the market. As described in chapter 4, the market for identity-relevant information on SNS does not function perfectly. There is especially no perfect information.

Both of these measures should be addressed in the policy development process. They can function as initial measures to base a discussion on and start the search for other preventive measures.

Another consideration to take into account is to incorporate the policy development process for privacy protection in Social Network Sites in Internet governance processes that are already taking place. The regulation of activities of Social Network Sites can very well be discussed within the light of technical standardization, Internet policy formulation and resource allocation. SNS use technical standards for the collection of information from other websites, and these standards could be designed in such a way that they restrict collection of information from secondary sources. The regulation of SNS fits well in discussions about net neutrality and the responsibility that content providers should bare for the information that is offered on their websites. Finally, the Internet is a world-wide network, and multiple organizations and governments need to be involved for effective regulation. As became clear from the CIPPIC complaint with the Canadian privacy officer, differences in privacy laws between countries make certain activities illegal in some countries and legal in others.

The following activities are part of the policy development process. This agenda identifies the various steps, however not in a chronological order:

- Identify important stakeholders and their interests and goals. The actors from the analysis in paragraph 4.3 and illustration 6 provide an initial list. A large portion of these actors will be attending the conference on 'Privacy in Social Network Sites' in Delft on October 23 and 24, 2008. The conference is a good starting point for

identifying more actors that have an interest in preventing privacy harms or have instruments and means to implement these measures. As mentioned before, the participation of SNS in this process is both crucial and uncertain.

- Design a policy development process, which should lead to detailed ideas on which preventive measures will be implemented and how these will be implemented. Because the incentives of many actors are incompatible, the process should be developed carefully to be successful. Following the proved method of De Bruijn, Ten Heuvelhof and In't Veld (2002), this process should protect the core values of the different parties to keep them committed, without losing the openness needed to come to solutions that are beneficial for all. Openness means that new actors should be able to join the process at any time. However, a strong emphasis should be on the substance as described in this thesis. The model to deconstruct privacy incidents acts as a model to identify exactly where preventive measures would be most effective. The focus on committing the actors to the process and its solutions should however not result in a very long process. The speed of the process is an important point of focus. The policy development process should start as soon as possible, to minimize any current privacy risks. Governments should have the primary responsibility for starting this policy process. Therefore, it is important that privacy issues in Social Network Sites get a higher position on the political agenda, as advocated by the position paper of ENISA (2007) and the Rome Memorandum of the IWGDPT (2008).

- Identify preventive measures. On the conference 'Privacy in Social Network Sites' a panel will discuss the capabilities that Privacy Enhancing Technologies (PETs) have to give users more control over their information and determine themselves how to restrict the access to the information. An initial long list of preventive measures is one of the expected outcomes of the conference. The presentation of Natali Helberger on the conference about legislation that influences the activities of SNS is relevant in this respect. Furthermore, the recommendations from ENISA (2007) and the International Working Party on Data Protection in Telecommunications (2008) for measures should be incorporated in this.

- Create consensus on preventive measures. The stakeholders that participate in the policy development process all have their own incentives. These incentives lead them to value specific preventive measures more than others. The process should foresee in defining criteria by which to assess the preventive measures and find ways to come to consensus on these criteria and their weights. A multi-criteria decision-support model, as used by Bots and Hulshof (1995) in the policy development for Dutch national healthcare, is an appropriate tool to come to consensus about the preventive measures that should be implemented. Although the interests of the various actors involved are opposed, decision-support models have recorded great successes for establishing consensus between actors with various interests.

- When consensus is reached on which preventive measures to implement in what order, the implementation process should be designed. For this process the same rules apply as for the development process, however it should be aimed at the implementation of regulations, laws and technical solutions to the problem. The implementation deserves attention in specific, because it encompasses the compliance

and monitoring of Social Network Sites with the regulations. As these have incentives to not comply, this should be designed thoroughly.

Despite the outline for a policy development process mentioned above, it has to be noted that policy makers do not always make rational decisions and that policy processes are rarely as structured as mentioned above. The case of Megan Meier, mentioned in paragraph 8.2, is a good example of this. The suicide of the 13-year-old girl shocked America and legislators were pushed to penalize the adult who played a role in the harassment of Megan. The analysis with the framework concludes that the harm was derived from the fact that the adult was able to contact Megan, and that she was able to send Megan harassing messages. Based on this reasoning, one would expect that lawmakers would come up with ways to prevent adults from contacting minors on SNS. However, prosecutors charged the adult with one count of conspiracy and three counts of computer breach, namely creating and using a fake profile. This does punish the adult, but does not protect minors from future harmful contact from adults on Social Network Sites. Furthermore, this creates a precedent for making pseudonymous use of SNS liable. With respect to restricting the access to identity-relevant information this is clearly undesirable. This example shows that policy makers under pressure do not always act in a way that would solve the problem most optimally.

10. Conclusions and recommendations

10.1. Conclusions

Users put their privacy at risk by uploading identity-relevant information to Social Network Sites. These websites allow users to self-disclose information by creating a profile, viewing the profiles of others and connecting to friends. Users write themselves into being in the online world, but at the same time provide SNS and potential wrongdoers with a wealth of information about them. This information, whether its first and last name or personal photos, describes details about their identity and is thus identity-relevant information. Identity-relevant information can be either referential, referring directly to a person (such as Social Security Number), or attributive, describing attributes to a person. Most privacy protection laws focus on the protection of referential data, thereby protecting a user against unwanted identification. However, many privacy threats for users of SNS stem from the widespread availability of attributive data, such as posting of information about a person by others and damage to one's reputation because of the dissemination if sensitive information.

There are four reasons why it is morally desirable to restrict the access to the identity-relevant information on Social Network Sites: the information could be used by wrongdoers to harm us, such as burglars who break into your house when they know you are at work because you have listed your job and home address. Or the substantial amount of attributive data is used to create a profile and the government uses this information to monitor political dissidents. This leads to the chilling of free speech, because of inequality in the information balance. Governments have more information about individuals than these individuals know or want. In addition, sensitive information about you, such as photos of your late-night beer-bingeing activities, ends up with exactly that group of people that you do not want to, such as your employers. The unwanted transgressing of information to these different groups or social spheres, called informational injustice, led to several outcries from SNS users. Finally, the online profiles on Social Network Sites become our digital counterparts, or digital persons. As users of SNS do not have much control over who adds information to their profile or to whom the information is disseminated, this prevents them for presenting their own moral identity and thus restricts their moral autonomy.

The trade-off between the will to disclose identity-relevant information to friends and the moral desirability to restrict access to this information (the privacy paradox) is visible when correlating the four reasons to restrict access to identity-relevant information with the three activities that SNS perform with the information: information collection, information processing and information dissemination. The resulting framework, as depicted in table 16, is suitable to deconstruct privacy incidents and come up with preventive measures to stop privacy threats from materializing into harmful incidents. For example, the mental inequality between an adult and a teenager that has contributed significantly to the extent of the harm to a teenager (see Megan Meier case in 8.2) could have been minimized by preventing the information-based harm from occurring. If the adult was not able to contact the teenager, or in other words if the contact information of the teenager was not disseminated from the SNS, the adult would not have been able to contact the teenager.

	Information-based Harm	Informational Inequality	Informational Injustice	Moral identity and autonomy
Information Collection				
Information Processing				
Information Dissemination				

Table 16 - The framework to deconstruct privacy incidents.

The biggest threat for users of SNS, in terms of negative impact and the amount of users that are affected, is government's usage of the information on SNS. Especially when a large part of it is collected from other (partnering) websites. This threat resembles the American government's Total Information Awareness program, except users are uploading information voluntarily to SNS. A less harmful threat, which does affect all users of SNS, is the minimal control they have over their own information. This is the green area depicted in illustration 9, which shows the 11 privacy risks for users of SNS. The minimal control over their identity-relevant information shows from the unwanted posting of information about users by others. It also shows from having no control over the secondary use of the information once uploaded, the unwanted dissemination to others and sharing and selling of their information with third parties. The probability that wrongdoers will obtain the information is a little smaller, although when this happens it can severely impact users. Wrongdoers could damage the reputation of users, stalk or bully them or use identity-relevant information to steal the user's identity.

Indeed, privacy incidents are complicated because many actors participate in the value chain of SNS, the incentives and interests of these actors contradict with those of the users, legal regulations do not (fully) apply and the design and topology of SNS stimulates the uploading and dissemination of as much identity-relevant information as possible. For example, Canada's privacy law is much more severe on the activities of SNS Facebook (Lawson 2008), than the tort laws in the USA. The OECD's Fair Information Practices are not used as regulatory norms by legislators and fined as such. Furthermore, the specific topology of SNS increase the speed of information dissemination, making it harder to prevent unwanted disclosure. This is also greatly facilitated by the stripping of contextual cues and the promotion of the disclosure of information by the default settings. Daniel Solove (2008a) characterizes the discrepancy between contextual cues in the real world and on Social Network Sites as follows:

> *"Social Network Sites are designed in a way that strips away certain kinds of contextual cues that ordinarily make people think twice [about uploading information]. If you would speak aloud the text in your web log in front of an auditorium of a 1000 people, you would be much more reticent, because you could see the people. This visualization of people would drastically alter the way you act. If you're sitting alone in your room in front of the computer, it's pretty hard to visualize 1000 people. It is hard to really grasp the full extent*

of it. [..] That makes it hard for people to see the consequences of putting information online."

The actors in the value chain have incentives to promote the free flow of identity-relevant information, so they can generate revenue from it. Users do not have to pay for most Social Network Sites, but the value generated from their identity-relevant information recoups the costs SNS incur. This usage of users' information leads to the privacy risks defined below.

Elaborate examples of the 11 privacy risks in illustration 9 are described in chapter 8 and deconstructed with the framework in table 16. These examples were not hard to find and the deconstruction leads to valuable results for designing preventive measures. The usability of the framework in table 16 to deconstruct incidents and come up with preventive measures increases its validity.

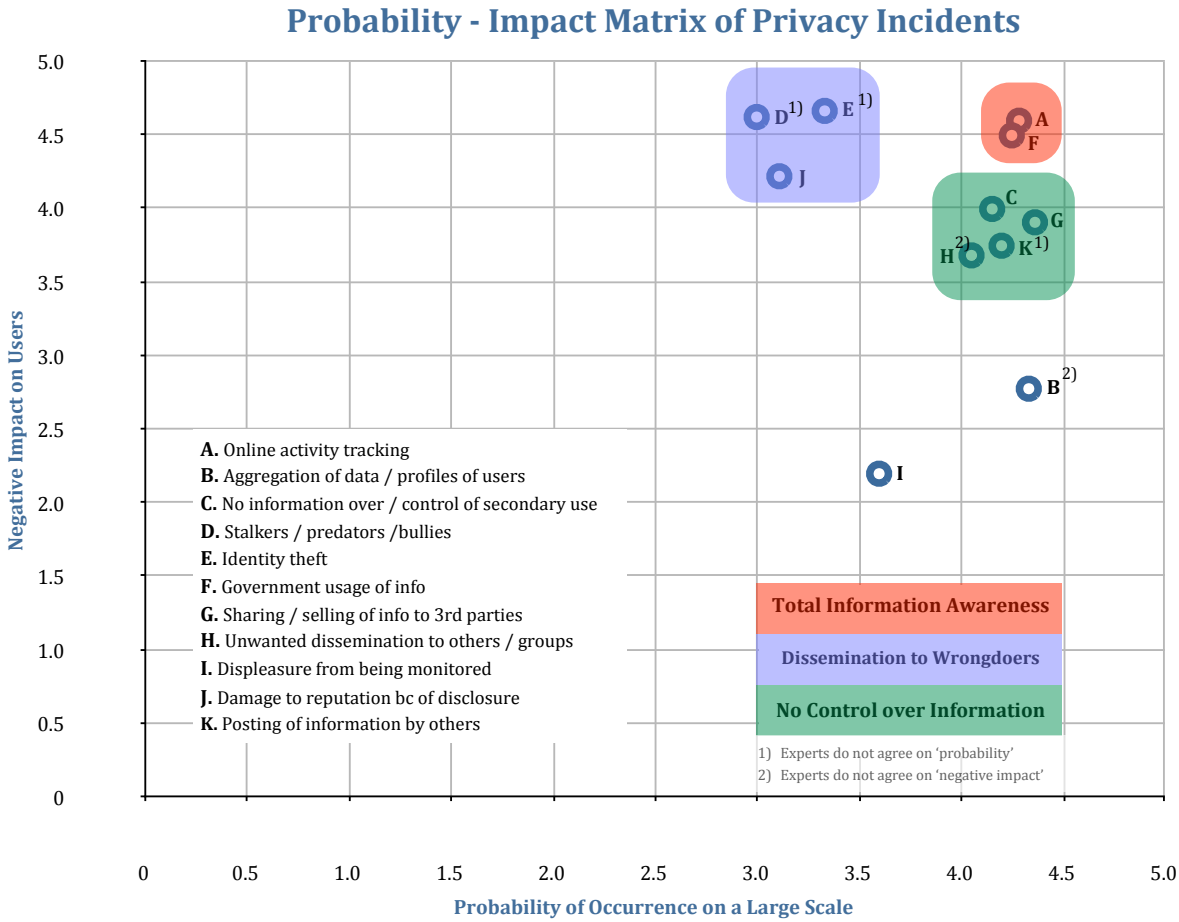


Illustration 12: Privacy incidents for users of Social Network Sites.

10.2. Recommendations for further research and research agenda

The research for this Master thesis was restricted to a limited time period and therefore not all issues could be dealt with at the same level of analysis.

It was very hard to obtain insight in how Social Network Sites and their partners create revenue and how much revenue and costs they incur for their different activities. In the first

place, SNS are not very open about what activities they execute with the collected information. Furthermore, they are (mostly) privately held companies, so it is hard to obtain information about their business. Indeed, as shown in paragraph 4.3 and table 1, the way that SNS generate revenue directly threatens the privacy of its users. Therefore, I recommend to conduct further research on the business activities that SNS perform, how they are generating revenue from these and what the associated costs are.

A better insight in the revenue streams and costs of SNS helps in defining a more specific utility function for Social Network Sites that accurately portrays their financial incentives. This increases the explanatory value of the game model in table 1 and the insight in possible preventive measures. Nonetheless, there is no quantitative function that describes the utility of the user of a SNS. This makes it difficult to identify to which degree the users of SNS accept privacy risks for the benefits they derive from disseminating their identity-relevant information to (a small group of) people. Methods and techniques are available to measure this utility curve and the trade-offs users make by for example interpreting quantitative surveys. Multivariate data analysis is suited to identify which factors specifically influence the user utility curve and how the pay-offs of this function are structured. Daniel Solove (2007a) mentions that the law should provide users with more control over their information, but the extent to which this complies with users' preferences is unknown. Thus I recommend to conduct a quantitative survey among users of Social Network Sites, and analyze this data to construct the users' utility curve. The more specific the utility functions of the users and the SNS are defined, the more explanatory power the game in table 1 has in terms of identifying trade-offs and preventive measures to influence the pay-off matrix.

The (perceived) benefits that users obtain from disclosing information on SNS is clear from the popularity of these websites. However, in several cases described in chapter 8 it showed that users might not want that information disseminates beyond certain groups or social spheres. Strahilevitz (2004) described a Social Network Theory of (electronic) privacy, which is very well suited to understand the expectations of privacy that users have. I recommend to adapt this theory specifically to Social Network Sites and use this analysis to identify preventive measures that allow users to derive benefits from the dissemination of their identity-relevant information without posing threats to privacy. Preventive measures could help users in keeping information within their own social circle (Solove 2007a). This could also lead to a more granular definition of friends.

Different types of this identity-relevant information morally ask for different types of protection. For example, more severe protection is required for what the EU defines as sensitive data. Leenes (2008) decomposes identifiability into four subcategories, that all should be regulated in a different way. He defines "L-, R-, C- and S-Identifiability. L-identifiability allows individuals to be targeted in the real world on the basis of the identifier, whereas this is not the case in the other three. R-type identifiability can be further decomposed in S-type, which is a technical kludge, and C-type identifiability which relates to the classification of individuals as being member of some set". I recommend to map these types on the conceptualization of identity-relevant information and Social Network Theory, to determine if and how different types of identity-relevant information should be differently regulated.

Not only do different types of identity-relevant information provide for different types of

identification, there are also various processes that can be used to identify or profile a user. One of the most complex and covert ways to identify users is the inference with (future) information from the user's social network. Users are part of social networks with coherent preferences, for example a network that support a band. These relationships can be used to identify users, even if they do not provide information on these subjects in their profile. Furthermore, information that is collected in the future can make an anonymous profile once of a sudden identifiable. Users of Social Network Sites leave snippets of identity-relevant information on various websites. This information can be combined and based on this identifying statements can be produced. It is recommended to look further into how individual members of Social Network Sites and their preferences can be identified by the preferences of their social network and what role future information can play in this inference.

In the reflections on the methodology in paragraph 9.1, I mentioned that external validation of the framework to deconstruct privacy incidents will contribute to the quality of the framework. It is recommended to test the framework in other areas where informational privacy is important, such as Electronic (online) Health Records or e-Commerce.

Finally, a Social Network Site can be conceptualized as a market place in which users sell their identity-relevant information to advertisers. The fact that most SNS are free, does not mean that users are not paying for it: the value generated from their information is used to make up for the costs of the SNS owner. In that respect, the SNS acts as a platform, bringing buyer and seller together. These markets are called two-sided markets (Rochet and Tirole 2004) and these market have specific characteristics with respect to the allocation of costs, revenues and externalities. Currently, there is a proposal to analyze SNS as two-sided markets in a paper building on the research of this thesis and the work of Milton Mueller in this field. This publication will involve Milton Mueller and myself.

The research agenda, not in a chronological order, is as follows:

- Scrutinize business activities of SNS and construct their utility function.
- Survey users of SNS and compose their utility function with the help of multivariate data analysis.
- Apply the Social Network Theory of privacy as coined by Strahilevitz (2004) to Social Network Sites and the identity-relevant information on those websites.
- Distinguish between several types of identifiability and analyze what this decomposition would mean for privacy threats and regulations in SNS.
- Identify methods for inference of identity with incomplete information or future information and assess how this could threaten the privacy of users of Social Network Sites.
- Apply the framework to other areas where informational privacy plays an important role.
- Analyze Social Network Sites with the theory of two-sided markets.

I. Glossary

Appropriation
Appropriation consists of using someone's else's identity-relevant information, in specific information that defines his likeness, to make a profit. In American tort law, this is translated into the appropriation tort. The appropriation tort creates a liability when someone uses the likeness of another for his own benefits (Allen 2006).

Assortativity of graph
The assortativity of a network measures whether these super nodes are more likely to connect with each other than with other users. Newman (2002) describes assortativity as the preferentially connection between vertices that have many connections (a high degree). He also mentions that social networks tend to show this behavior and calls them assortatively mixed, while technological and biological networks tend to be disassortatively mixed.

Attributive data
Data that describes attributes to people and in is not descriptive enough to uniquely identify a person. However, aggregated attributive data could identify a person. See (Van den Hoven 2007)

Architecture of control and vulnerability
Solove (2004) defines an architecture of control as "an architecture that structures power, a regulatory framework that governs how information is disseminated, collected and networked". This skewed power balance leaves the less powerful in a vulnerable position, thus the architecture of vulnerability, a term that Solove also uses. In his newer book, 'The digital person' (2006), he describes how profiles of people stored in databases that are used to make decisions about them, put them in a vulnerable position, because it is unclear what the decisions are based on and they can not influence this process. For an example, see the Herring case as described in paragraph 8.1.

Average shortest path
The average shortest path is the mean of the set of shortest paths between any two users, which can be measured.

Blocked exchange
A way to prevent dissemination of information to multiple social spheres is using 'blocked exchanges' that separate information from one sphere to another. Van den Hoven (2007) mentions that a fine-grained authorization matrix can serve as a blocked exchange between the different spheres.

Clustering coefficient
The clustering coefficient determines how well friends of you are connected to each other. Alber-Laszlo and Barabasi (2003) define the clustering coefficient mathematically as the actual numbers of edges a vertice has divided by the number of edges the same vertice could maximally have.

Content-Based Image Retrieval (CBIR)
Content-Based Image Retrieval is an emerging technology that can match features, such as an identifying aspect of a room (a painting) to a database with location information, increasing the possibilities for identifying users (ENISA 2007).

Cookie
Cookies are small text files that websites place on a users computer. These small text files can be retrieved by the webserver later and are therefore referred to as cookies. Social Network Sites use cookies to store your login credentials. Chester (2008) mentions that marketing profiling companies use cookies to set your interests. Cookies are divided in several areas, each one specifying an interest area. Whenever you visit for example a website that is about cars, the interest cookie would be updated in the area about cars.

Cross-site Scripting (XSS)
Including scripts that are running on one website to manipulate data on other websites. The scripts on the first website are not governed or monitored by the second website. For example, on many SNS users can post information on their profile in HTML. If the SNS allows javascript, a malicious javascript from another website can be run on the first one, compromising the security of the first website.

Data aggregation
The aggregation of data entails bringing data together from disparate data sources into one profile. Information that separately would not have a distinct meaning, might become meaningful or interesting for advertisers once it is combined. Solove (2006) mentions that the whole becomes bigger than the parts and can reveal new facts about a person that he did not expect to be known. He also mentions that people deliberately spread little pieces of information about themselves throughout their daily activities, because they do not expect all this information to be combined a create a full picture of them.

Degree
The degree is the number of connections, or number of edges, that a node or vertice has. The average degree is a valuable measurement for assessing the connectivity in a network.

Digital Immigrants
Marc Prensky (2005) defines 'Digital Natives' as "today's students [,] [n[ative speakers of technology, fluent in the digital language of computers, video games, and the Internet." He contrasts them with 'Digital Immigrants', the people that adopt and have to learn every new technology.

Edge
An edge is a connection between to nodes or vertices. The term 'edge' is used in graph theory, in the ICT science it is more commonly known as a link.

Friends list
List of all contacts that have a direct connection with the SNS user. Social Network Sites and researchers commonly refer to these contact as friends, although they technically do not have to be friends in real life. Friends lists typically represent who users care about and perform the function of an imaginary audience (boyd 2007).

Graphical User Interface (GUI)
A Graphical User Interface, commonly abbreviated as GUI, is the graphical representation of computer software that is presented to the user. The user navigates the GUI and interacts with the various screens that make up the GUI.

Informed consent
Informed consent means that the subject of the collection, processing and dissemination of his information should be fully informed by the data processor and fully consent. The definition

builds on two concepts: the subject should be fully informed and the subject should give an explicit consent. Privacy activists have complained about the actions that SNS take to inform users. According to Acquisti and Gross (2005) more than 84% of the users does not read the privacy policy, which should inform them about the collection, processing and dissemination of information. A good example of the dubious informing of users is the Beacon example (see paragraph 3.5). Solove (2008) focuses more on the second part when saying that the informed consent that most SNS require does not leave much freedom for the users: "[They are] given a take-it or leave-it choice."

Least authority / least privileges principle
The principle of least authority states that a function should have the least amount if authority needed to fulfill its functions (Soghoian 2008b). For example, with respect to accessing databases, a function would only be able to query the data it specifically needs for its operations.

Limited privacy
The concept of limited privacy implies that when an individual reveals private information about herself to one or more people, she may retain a reasonable expectation that the recipients of the information won't disseminate it further (Strahilevitz 2005).

Mediated and unmediated publics
Mediated publics (boyd 2007) are publics that have to a certain extent control over what happens with the information that is conveyed through them. Therefore these mediated publics, such as newspapers or television channels, have specific characteristics. For example, newspapers have a specific public and whatever is written in them can be retained for a long period. Unmediated technologies, such as a public spot on the street, do not have these characteristics.

Networked publics
When mediated publics ported to the World Wide Web and became connected to each other, they became online mediated environments. The online (networked) mediated environments have the following characteristics: persistence (what you say sticks around), searchability of the information, replicability (digital bits are easy to copy) and invisible audiences (you are not able to fully determine who you are communicating with) (boyd 2007).

Opt-in
Opt-in requires a notification to the user and his explicit consent to collect, process or disseminate information. The opt-in requires affirmative action by the data subject for the collection, processing or dissemination of the information (Rotenberg 2001), in contrast to the opt-out option.

Opt-out
Opt-out assumes consent unless an individual affirmatively indicates a preference for not sharing the information. However, Solove (Solove 2006) points out that this doesn't imply informed consent and could lead to insecurity or unawareness of secondary use. The problem with opt-out is that it has lower transaction costs for the companies processing the data than the opt-in option (Rotenberg 2001).

Phishing
Phishing attacks involve the creation of a fake website or email, which resembles a real-world website or email (social engineering), to abuse the users trust in this website and inquire for

logins and passwords. Spear phishing is a highly personalized form of phishing, directed at one person or a group of person, thereby increasing the trust.

Preferential attachment
Nodes with specific features, such as a high degree, sometimes connect to other nodes with the same features more likely than with other nodes. This is called preferential attachment and it could explain why information travels at specific speeds through a network. The clustering coefficient measures if nodes with a high degree preferential connect to each other.

Privacy paradox
Users disseminate identity-relevant information to others, thereby increasing risks of privacy harms. The paradox lies in the fact that users want to disseminate information, thereby increasing risk, but still expect that the information stays private or within a specific group of people. Barnes (2006) coined the term 'privacy paradox' and describes it as "young people are pouring their minds, if not their hearts, into cyberspace", but assume a certain amount of control over who sees this outpouring.

Purpose specification
The use limitation principle is one of the eight Fair Information Practices as defined by the OECD in 1980. It states that "The purposes for which personal data are collected should be specified not later than at the time of data collection and the subsequent use limited to the fulfillment of those purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose" (Gellman 2008).

Referential data
Data that is about one person and uniquely identifies this person (Van den Hoven 2007).

Secrecy paradigm
The secrecy paradigm relies on the concept of private or secret information to identify privacy harms. If information is available to the public, it is not deemed private and therefore usage of the information does not cause harms in terms of privacy. The secrecy paradigm focuses on breached confidentiality, harmed reputation and unwanted publicity. These harms are however not the central problems of databases (Solove 2004).

Secondary use of information
The data that Social Network Sites collect should be collected for a specific purpose and their users should be notified of this purpose, as stated by the OECD's use limitation principle (Gellman 2008). However, in practice this information is stored in databases and used for other purposes that benefit the SNS. This secondary use of information is not in compliance with the OECD's Fair Information Practices if the informed consent of the user is not obtained. If no informed consent is obtained, it directly contradicts with the Fair Information Practice's use limitation principle (Gellman 2008).

Security through obscurity
Principle in security engineering that relies on the secrecy of potential security flaws to provide security (FIND REFERENCE) (Wikipedia 2008b). SNS Facebook is said to deploy this principle while offering third-party application providers access to all their users' information (Soghoian 2008a).

Social Network Site
A Social Network Sites is a websites that:

1. Primarily allows users to divulge information (self-disclosure) by constructing a public or semi-public profile (ENISA 2007), which contains identity-relevant information (referential as well as attributive), within a bounded system, and which information is stored by the SNS in databases.

2. Enables users to articulate a list of other users (friends list)[49], with whom they share a connection, and make this list visible to other users to facilitate contact between users and interact in various ways with these users (ENISA 2007), sometimes by making use of the proprietary algorithms that SNS provide.

3. Enables users to view and traverse their list of connections and those made by others within the Social Network Site, by means of a graphical user interface on top of the database of profiles, and enable users to: connect to one another with bilateral agreement; define the social relationships they have with their friends (ENISA 2007) leave public testimonials or comments (boyd 2007); or create and join communities that share the same interests and activities (Fu et al 2007).

4. Is characterized by membership rules with their own community standards and sustained social interaction between their members (Rothaermel 2001 based on Lawrence 1995 and Karp et al. 1977).

Social Network Site aggregator
Social Network Site aggregators are websites that collect, process and disseminate information from multiple platforms on just one website. Aggregators like Rapleaf.com and Wieowie.nl gather data from the networks whenever a user submits a query for a person. Websites like Snag and Profilelinker combine the management of several Social Network Sites by providing one central access point, thereby multiplying the vulnerabilities of the accounts by giving access and read / write permissions to multiple accounts based on weak authentication (ENISA 2007).

Third party doctrine
American doctrine that states that if information is possessed or known by third parties, the, for purpose of the fourth amendment and individual lacks the reasonable expectation of privacy in the information {Citation}. Although this doctrine is based on the American Fourth Amendment, which protects against unreasonable searches and seizures by government officials, the principle argumentation is relevant.

Unconscionable contract
An unconscionable contract is a specific form of deceptive and unfair contract that has been described as "one which no man in his senses, not under delusion, would make ... and which no fair and honest man would accept" (Leng 2006). Leng (2006) specifically looks at the unconscionable contract in the light of eCommerce, and finds that the doctrine of unilateral mistake is well settled in the United States, Singapore and Australia. With other words, if one party can prove it is deceived, the contract is void.

Use limitation
The use limitation principle is one of the eight Fair Information Practices as defined by the OECD in 1980. It states that "Personal data should not be disclosed, made available or otherwise used for purposes other than those specified in accordance with [the Purpose

[49] Refer to footnote 4 on SNS 'friends' and friends' list.

Specification Principle] except: a) with the consent of the data subject; or b) by the authority of law (Gellman 2008).

Value chain
The value chain is the set of activities that a firm employs to create value. Porter (1985) initially identified the following in the value chain: inbound logistics, operations, outbound logistics, marketing & sales and service. The digitization of information and the possibility to offer services online has changed the value chain significantly. Outbound logistics are replaced by the dissemination of digital information via the Internet and inbound logistics contain user-generated content. Furthermore, most of the operations in Social Network Sites are executed by algorithms that for example determine which information is disseminated to which person at what time.

Vertex
A vertex is an entity that other entities can connect to (Gross 2006).Vertices are known in ICT science as nodes. Social Network Sites can be modeled as graphs, in which the users are vertices and the links between users edges that connect the vertices.

Web beacon
Web beacons consist of a small string of code that represents a graphic image request on a web page or e-mail (Krunic et al. 2006). With these small strings of code, information can be collected from one websites and passed on to another website.

II. References

2008d. "Hyves moet zich verantwoordelijker opstellen". Mijn Kind Online (www.mijnkindonline.nl): *. Available at: http://mijnkindonline.web-log.nl/mijnkindonline/2008/04/hyves-moet-zich.html [Accessed August 23, 2008].

2002. Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications). Available at: http://eur-lex.europa.eu/smartapi/cgi/sga_doc?smartapi!celexapi!prod!CELEXnumdoc&lg=en&numdoc=32002L0058&model=guichett [Accessed September 5, 2008].

1995. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data . Available at: http://europa.eu.int/eur-lex/lex/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:HTML [Accessed September 3, 2008].

2008. Facebook 'violates privacy laws'. *BBC*. Available at: http://news.bbc.co.uk/2/hi/technology/7428833.stm [Accessed June 2, 2008].

2006. *Internet Governance and Regulation: The Future of the Internet - and How to Stop It*, Oxford internet institute webcast . Available at: http://webcast.oii.ox.ac.uk/?view=Webcast&ID=20060411_141 [Accessed March 21, 2008].

2008b. Protest tegen de verandering - Hyves.nl. Available at: http://tegendeverandering.hyves.nl/ [Accessed August 23, 2008].

1989. Requirements for Internet Hosts -- Communication Layers. Available at: http://tools.ietf.org/html/rfc1122.

2008a. The 'MySpace suicide' trial. *The Los Angeles Times*. Available at: http://www.latimes.com/news/opinion/la-ed-myspace19-2008may19,0,7075638.story [Accessed June 4, 2008].

Acquisti & Gross, 2006. Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook. In Privacy Enhancing Technologies. pp. 36-58. Available at: http://dx.doi.org/10.1007/11957454_3 [Accessed April 1, 2008].

Aleman-Meza, B. et al., 2006. Semantic analytics on social networks: experiences in addressing the problem of conflict of interest detection. In Proceedings of the 15th international conference on World Wide Web. Edinburgh, Scotland: ACM, pp. 407-416. Available at: http://portal.acm.org/citation.cfm?id=1135777.1135838 [Accessed August 23, 2008].

Allen, A., 2007. *Privacy law and society*, [St. Paul Minn.]: Thomson/West.

Allen, C., 2006. Collective Choice: Using Five-Star Rating Systems. Available at: http://www.skotos.net/articles/TTnT_/TTnT_194.phtml.

ANP, 2008. Rechter: Afgeschermde Hyves ook openbaar. NRC Handelsblad. Available at: http://www.nrc.nl/media/article1946030.ece/Rechter_Afgeschermde_Hyves_ook_openbaar [Accessed August 22, 2008].

Article 29 Data Protection Working Party, 2007. *10th Annual Report on the situation regarding the protection of individuals with regard to the processing of personal data in the European Union and in third countries - covering the year 2006.*, Available at: http://ec.europa.eu/justice_home/fsj/privacy/workinggroup/annual_reports_en.htm [Accessed May 9, 2008].

Axten, S., 2008. Facebook data breach. Available at: http://www.oag.state.md.us/idtheft/Breach%20Notices/ITU-153491.pdf [Accessed August 7, 2008].

Barabasi, A.L. & Crandall, R.E., 2003. Linked: The New Science of Networks. American Journal of Physics, 71, 409.

Barnes, S.B., 2006. A privacy paradox: Social networking in the United States. *First Monday*, 11(9).

Bayard, S., 2008. Max Mosley's S&M Party Not A Matter of Legitimate Public Concern, Says English Court. Citizen Media Law Project. Available at: http://www.citmedialaw.org/blog/2008/max-mosleys-sm-party-not-matter-legitimate-public-concern-says-english-court [Accessed August 17, 2008].

Benassi, P., 1999. TRUSTe: an online privacy seal program. Commun. ACM, 42(2), 56-59.

Berliner Beauftragter für Datenschutz und Informationsfreiheit, Europa / International / International Working Group on Data Protection in Telecommunications (IWGDPT). Available at: http://www.datenschutz-berlin.de/content/Europa+%252F+International/International+Working+Group+on+Data+Protection+in+Telecommunications++%28IWGDPT%29 [Accessed September 3, 2008].

Bots, P.W. & Hulshof, J.A., 1995. *Applying Multi-Criteria Group Decision Support to Health Policy Formulation*, Hong Kong.

Bots, P.W. & Hulshof, J.A., 2000. Designing multi-criteria decision analysis processes for priority setting in health policy. *Journal of Multi-Criteria Decision Analysis*, 9(1-3), 56-75.

boyd, d., 2006a. Facebook's "Privacy Trainwreck": Exposure, Invasion, and Drama. Available at: http://www.danah.org/papers/FacebookAndPrivacy.html [Accessed July 27, 2008].

boyd, d., 2004. Friendster and Publicly Articulated Social Networks. *Conference on Human Factors and Computing Systems (CHI 2004), Vienna, Austria, April*, 24-29.

boyd, d., 2006. *Identity production in a networked culture: Why youth heart MySpace*,

boyd, d., 2007a. Social network sites: Public, private, or what? *The Knowledge Tree: An e-Journal of Learning Innovation*.

boyd, d., 2007b. Why youth (heart) social network sites: The role of networked publics in teenage social life. *Identity. MacArthur Foundation http://www.danah.org/papers/[March 15, 2007]*.

boyd, d. & Heer, J., 2006b. *Profiles as Conversation: Networked Identity Performance on Friendster*,

Bright, A., 2008. Publication of Private Facts. Citizen Media Law Project. Available at: http://www.citmedialaw.org/subject-area/publication-private-facts [Accessed August 17, 2008].

de Bruijn, H., ten Heuvelhof, E. & Veld, R.I., 2002. Process Management: Why Project Management Fails in Complex Decision Making Processes, Kluwer Academic.

CA Security Advisor Research, 2007. Facebook SocialAds – Going Too Far? . Available at: http://community.ca.com/blogs/securityadvisor/archive/2007/11/29/facebook-socialads-going-too-far.aspx [Accessed April 8, 2008].

CA Security Advisor Research , 2007a. Facebook's Misrepresentation of Beacon's Threat to Privacy: Tracking users who opt out or are not logged in. Available at: http://community.ca.com/blogs/securityadvisor/archive/2007/11/29/facebook-s-misrepresentation-of-beacon-s-threat-to-privacy-tracking-users-who-opt-out-or-are-not-logged-in.aspx [Accessed April 8, 2008].

Cashmore, P., 2006. Facebook To Offer Credit Cards. Available at: http://mashable.com/2006/08/15/facebook-to-offer-credit-cards/ [Accessed August 16, 2008].

Chester, J., 2007. The Brandwashing of America: Marketing and Micro-persuasion in the Digital Era. In Digital Destiny: New Media and the Future of Democracy. New Press: Distributed by WW Norton.

College Bescherming Persoonsgegevens, 2007. Jaarverslag 2007,

Cox, P., 2002. Directive 2002/58/EC of the European Parliament and of the Council. Available at: http://eur-lex.europa.eu/smartapi/cgi/sga_doc?smartapi!celexapi!prod!CELEXnumdoc&numdoc=32002L0058&model=guichett&lg=en [Accessed May 24, 2008].

Daily Kos, 2007. Delete My Bleeping Account, Facebook! Available at: http://www.dailykos.com/storyonly/2007/12/25/18521/907 [Accessed April 9, 2008].

Edwards, L. & Brown, I., Data Control and Social Networking: Irreconcilable Ideas? SSRN eLibrary. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1148732#PaperDownload [Accessed July 10, 2008].

Elatrash, S., 2007. The wrong profile. A Mirror writer discovers someone has been impersonating him on Facebook . Available at: http://www.montrealmirror.com/2007/041207/news2.html [Accessed August 17, 2008].

Ellison, N.B. & boyd, d., 2007. Social Network Sites: Definition, History, and Scholarship.

Emily Harding, 2005. Electronic Law Journals - JILT 1996 (3) - van den Hoven & Cushman. Available at: http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/1996_3/hoven [Accessed March 17, 2008].

ENISA, 2007. ENISA Position Paper No.1 Security Issues and Recommendations for Online Social Networks.

EPIC, 2008. Facebook Privacy Page. Available at: http://epic.org/privacy/facebook/default.html [Accessed April 4, 2008].

EPIC, 2008c. Herring v. US Page. Available at: http://epic.org/privacy/herring/ [Accessed August 23, 2008].

EPIC, 2008b. Information Fusion Centers and Privacy. Available at: http://epic.org/privacy/fusion/ [Accessed August 16, 2008].

EPIC , 2008a. Social Security Number (SSN) Privacy Page. Available at: http://epic.org/privacy/ssn/ [Accessed April 25, 2008].

EPIC , 2005. Terrorism (Total) Information Awareness Page. Available at: http://epic.org/privacy/profiling/tia/ [Accessed August 16, 2008].

Facebook, 2008. Actions from External Websites (Beacon). Available at: http://www.facebook.com/help.php?page=56 [Accessed April 8, 2008].

Facebook , 2007c. Developer Terms of Service. Available at: http://developers.facebook.com/terms.php [Accessed April 8, 2008].

Facebook, 2007a. Privacy Policy. Available at: http://www.facebook.com/policy.php [Accessed April 15, 2008].

Facebook, 2007b. Terms of Use. Available at: http://www.facebook.com/terms.php [Accessed May 8, 2008].

FEDERATION OF EUROPEAN DIRECT MARKETING , EUROPEAN CODE OF PRACTICE FOR THE USE OF PERSONAL DATA IN DIRECT MARKETING .

Felt, A., 2008a. Interview with Adrienne Felt about her work on 'Third party applications for Facebook'.

Felt, A., 2008b. Privacy Protection for Social Networking APIs. *Unpublished, available from Adrienne Felt*.

Fu, F., Wang, L. & Liu, L., 2007 Empirical analysis of online social networks in the age of Web 2.0. Physica A: Statistical Mechanistics and its Applications. Available at: In Press, Corrected Proof.

Fu, F. et al., 2007a. Social dilemmas in an online social network: The structure and evolution of cooperation. Physics Letters A, 371(1-2), 58-64.

Gavison, R., 1980. Privacy and the Limits of Law. Yale Law Journal, 421.

Gedragscode Verwerking Persoonsgegevens Financiële Instellingen.

Gellman, B., 2008. FAIR INFORMATION PRACTICES: A Basic History. Available at: http://bobgellman.com/rg-docs/rg-FIPshistory.pdf.

Google, 2007. Urchin Web Analytics - Urchin Traffic Monitor (UTM). Available at: http://www.google.com/support/urchin45/bin/answer.py?answer=28710 [Accessed September 5, 2008].

Gosden, E., 2007. Oxford students' trial by Facebook . The Guardian. Available at: http://www.guardian.co.uk/media/2007/jul/17/digitalmedia.highereducation [Accessed August 22, 2008].

Grimmelman, J., The Laboratorium: Facebook and the VPPA: Uh-Oh. Available at: http://laboratorium.net/archive/2007/12/10/facebook_and_the_vppa_uhoh [Accessed April 9, 2008].

Gross, D., 2006. $1 billion for Facebook! LOL! Is the social-networking boom a replay of the '90s dotcom bubble?

Gross, R., Acquisti, A. & Heinz III, H.J., 2005. Information revelation and privacy in online social networks. Proceedings of the 2005 ACM workshop on Privacy in the electronic society, 71-80.

Gunther, M., 2006. News Corp. (hearts) MySpace. Fortune. Available at: http://money.cnn.com/2006/03/28/technology/pluggedin_fortune/index.htm [Accessed September 11, 2008].

Hargittai, E., 2007. Whose Space? Differences Among Users and Non-Users of Social Network Sites. *Journal of Computer-Mediated Communication*, 13(1), 276-297.

Healy, G., 2003. Beware of Total Information Awareness. CATO Institute. Available at: http://www.cato.org/pub_display.php?pub_id=2959 [Accessed August 16, 2008].

Hoegg, R. et al., 2006. Overview of business models for Web 2.0 communities. *Proceedings of GeNeMe*, 23-37.

Hoegg, R. & Stanoevska-Slabeva, K., 2005. Towards Guidelines for the Design of Mobile Services. *Proceedings of the ECIS 2005 conference, June*.

Hollywood, J., 2004. Out of the Ordinary: Finding Hidden Threats by Analyzing Unusual Behavior, Rand Corp.

van Den Hoven, J. & Lokhorst, G., 2002. Deontic Logic and Computer-Supported Computer Ethics. *Metaphilosophy*, 33(3), 376-386.

van den Hoven, J. & Manders-Huits, N., 2006. Identiteitsmanagement en morele identificatie. *Algemeen Nederlands tijdschrift voor wijsbegeerte*, 98(2).

van den Hoven, J., 2007. Information Technology, Privacy and The Protection of Personal Data. In J. Van den Hoven, ed. *Information Technology, Privacy and The Protection of Personal Data*. Cambridge, UK ; New York : Cambridge: University Press, pp. 462-494. Available at: http://lccn.loc.gov/2007016850.

van den Hoven, J., 2008. Information Technology, Privacy and the Protection of Personal Data.

van den Hoven, J. et al., 2008. Kick-off meeting graduation project. Available at: See appendices.

van den Hoven, J. et al., 2008a. Mid-term meeting graduation project. Available at: See appendices.

van den Hoven, M.J., 1997. Privacy and the varieties of moral wrong-doing in an information age. *ACM SIGCAS Computers and Society*, 27(3), 33-37.

International Working Group on Data Protection in Telecommunications, 2008. Report and Guidance on Privacy in Social Network Services - "Rome Memorandum" - . Available at: http://www.datenschutz-berlin.de/attachments/461/WP_social_network_services.pdf?1208438491 [Accessed August 11, 2008].

Jackson, H., 2008. Bebo party story is fake--lawsuit is not. News - Digital Media - CNET News. Available at: http://news.cnet.com/8301-1023_3-9989415-93.html [Accessed August 22, 2008].

Jones, H. & Soltren, J.H., 2005. Facebook: Threats to Privacy. *Project MAC: MIT Project on Mathematics and Computing*.

Karp, D.A., Stone, G.P. & Yoels, W.C., 1977. Being Urban: A Social Psychological View of City Life, Heath.

Kharif, O., 2007. Social-Networking Sites Open Up. *BusinessWeek*.

Klosek, J., 2000. Data Privacy in the Information Age, Quorum Books.

Knight, W., 2007. What are your embarassing Facebook pics? New Scientist Technology Blog. Available at: http://www.newscientist.com/blog/technology/2007/07/what-are-your-embarassing-facebook-pics.html [Accessed August 22, 2008].

Krunić, T. et al., 2006. Web Design Curriculum and Syllabus Based on Web Design Practice and Students' Prior Knowledge. JOURNAL OF INFORMATION TECHNOLOGY EDUCATION, 5, 317-335.

Kumar, R., Novak, J. & Tomkins, A., 2006. Structure and evolution of online social networks. In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. Philadelphia, PA, USA: ACM, pp. 611-617. Available at: http://portal.acm.org/citation.cfm?id=1150476 [Accessed April 1, 2008].

Kuner, C., 2007. *European Data Protection Law: Corporate Regulation and Compliance* 2nd ed., Oxford University Press, USA.

Lawrence, T.B., 1995. Power and resources in an organizational community. Academy of Management Best Papers Proceedings, 251-5.

Lawson, P., 2008. PIPEDA Complaint: Facebook. Available at: http://www.cippic.ca/cippic-news/ [Accessed June 4, 2008].

Leenes, R., 2008. Do You Know Me? Decomposing Identifiability. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1084878.

Leng, T.K., 2006. Legal effects of input errors in eContracting. Computer Law & Security Report, 22(2), 157-164.

Lyon, D., 1993. An electronic panopticon? A sociological critique of surveillance theory. Sociological review(Keele), 41(4), 653-678.

Malik, O., 2007. Classmates: Cashing In On Social Networking Hype. Available at: http://gigaom.com/2007/08/13/classmates-cashing-in-on-social-networking-hype/ [Accessed September 11, 2008].

Manders-Huits, N. & van den Hoven, J., Moral identification in Identity Management Systems.

Marks, P., 2006. Pentagon sets its sights on social networking websites. New Scientist. Available at: http://technology.newscientist.com/channel/tech/mg19025556.200-pentagon-sets-its-sights-on-social-networking-websites.html [Accessed August 23, 2008].

Martin, E.C., False Light. Stanford Netlaw. Available at: http://netlaw.samford.edu/Martin/AdvancedTorts/falselight.htm [Accessed August 17, 2008].

Matyszczyk, C., 2008. The Bebo party case. A ball of confusion. Technically Incorrect - CNET News. Available at: http://news.cnet.com/8301-17852_3-9989761-71.html [Accessed August 22, 2008].

McCarthy, C., 2008. AOL buys social network Bebo for $850 million. CNET News.com. Available at: http://news.cnet.com/8301-13577_3-9893014-36.html [Accessed September 11, 2008].

McGeveran, W., 2007. Facebook Inserting Users Into Ads. Info/Law. Available at: http://blogs.law.harvard.edu/infolaw/2007/11/08/facebook-social-ads/ [Accessed August 7, 2008].

McGeveran, W., 2007a. More Thoughts on Facebook's "Social Ads". Info/Law. Available at: http://blogs.law.harvard.edu/infolaw/2007/11/09/more-thoughts-on-facebooks-social-ads/ [Accessed August 7, 2008].

Mills, E., 2008. Psychological profiling on the Web. CNET News.com. Available at: http://news.cnet.com/8301-1009_3-10022743-83.html [Accessed August 24, 2008].

National Center for Transgender Equality: , 2007. Voting. Available at: http://www.nctequality.org/take_action/voting.html [Accessed August 23, 2008].

Newman, M.E.J., 2002. Assortative Mixing in Networks. Physical Review Letters, 89(20), 208701.

Nielsen, 2008. nielsen-online-top-10-social-networking-sites-us-february-2008.jpg. Available at: http://www.marketingcharts.com/interactive/top-10-us-social-network-and-blog-site-rankings-issued-for-feb-3851/nielsen-online-top-10-social-networking-sites-us-february-2008jpg/ [Accessed March 18, 2008].

Nissenbaum, H., Privacy as Contextual Integrity. *SSRN eLibrary*. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=534622 [Accessed June 3, 2008].

Noguchi, Y., 2006. Saying It 'Messed Up,' Facebook Modifies Controversial Feature. The Washington Post, D01.

Novum/ANP, 2008. Politie spoort criminelen op via Hyves. Trouw. Available at: http://www.trouw.nl/laatstenieuws/ln_binnenland/article993177.ece/Politie_spoort_criminelen_op_via_Hyves [Accessed August 25, 2008].

OECD, 1980. *Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*, Available at: http://www.oecd.org/document/18/0,2340,es_2649_34255_1815186_1_1_1_1,00.html [Accessed April 24, 2008].

Out-law.com, 2006. US government funds social network snooping. The Register. Available at: http://www.theregister.co.uk/2006/07/03/us_govt_funds_online_snooping/ [Accessed August 17, 2008].

Paget, F., 2008. You have to pay for quality. Computer Security Research - McAfee Avert Labs Blog. Available at: http://www.avertlabs.com/research/blog/index.php/2008/05/07/you-have-to-pay-for-quality/ [Accessed September 16, 2008].

Porter, M.E., 1985. Competitive advantage, Free Press New York.

Prensky, M., 2005. Listen to the Natives . Available at: http://www.ascd.org/authors/ed_lead/el200512_prensky.html [Accessed September 3, 2008].

Privacy Coalition, About the Privacy Coalition. Available at: http://privacycoalition.org/about.php [Accessed August 14, 2008].

Quinn, G., The Right of Publicity. IPWatchdog.com. Available at: http://www.ipwatchdog.com/trademark/the-right-of-publicity/ [Accessed August 17, 2008].

Radio Online, 2008. Heisa om Hyves. Available at: http://www.radio-online.nl/ [Accessed May 5, 2008].

Riphagen, D., 2008h. David Riphagen quoted in Dutch newspaper NRC Handelsblad. Privacy in Social Network Sites. Available at: http://privacyinsocialnetworksites.wordpress.com/2008/08/07/david-riphagen-quoted-in-dutch-newspaper-nrc-handelsblad/ [Accessed August 7, 2008].

Riphagen, D., 2008a. Do Social Networking Sites Have Small-World Effects? *Unpublished. available from author.*, 8.

Riphagen, D., 2008g. Facebook and the DPPA: Uh-Oh. Privacy in Social Network Sites. Available at: http://privacyinsocialnetworksites.wordpress.com/2008/09/05/facebook-and-the-dppa-uh-ohfacebook-and-the-dppa-uh-oh/ [Accessed August 5, 2008].

Riphagen, D., 2008b. Facebook Case Study.

Riphagen, D., 2008i. Not verifying lies on Facebook, media get sued for breach of privacy. Privacy in Social Network Sites. Available at: http://privacyinsocialnetworksites.wordpress.com/2008/07/16/not-verifying-lies-on-facebook-media-get-sued-for-breach-of-privacy/ [Accessed August 22, 2008].

Riphagen, D., 2008f. Prosecutors and lawyers turn to Social Network Sites. Privacy in Social Network Sites. Available at: http://privacyinsocialnetworksites.wordpress.com/2008/07/21/prosecutors-and-lawyers-turn-to-social-network-sites/ [Accessed August 5, 2008].

Riphagen, D., 2008c. Social Network Sites and social spheres? . Privacy in Social Network Sites. Available at: http://privacyinsocialnetworksites.wordpress.com/2008/07/09/social-network-sites-and-social-spheres/ [Accessed July 28, 2008].

Riphagen, D., 2008d. Social Network Sites and Third Parties.

Riphagen, D., 2008e. Wie is het online publiek: webloggers aan de schandpaal? *Geencommentaar.nl*.

Rochet, J.C. & Tirole, J., 2004. Two-Sided Markets: An Overview. Institut d'Economie Industrielle working paper.

Roschke, G., 2008. Interview with Guilherme Roschke about his work on Facebook and Social Network Sites at EPIC.

Rosenbush, B.S., 2005. News Corp.'s Place in MySpace. BusinessWeek: Technology. Available at: http://www.businessweek.com/technology/content/jul2005/tc20050719_5427_tc119.htm [Accessed September 11, 2008].

Rotenberg, M., 2001. Fair Information Practices and the Architecture of Privacy (What Larry Doesn't Get). *Stanford Technology Law Review*, 1.

Rotenberg, M., Information Privacy Law 2008. Available at: http://epic.org/misc/gulc/ [Accessed May 8, 2008].

Rotenberg, M., 2005. *Privacy Law Sourcebook 2004: United States Law, International Law, and Recent Developments*, Epic.

Rothaermel, F.T. & Sugiyama, S., 2001. Virtual internet communities and commercial success: individual and community-level theory grounded in the atypical case of TimeZone.com. Journal of Management, 27(3), 297-312.

Ruhami, A., 2006. Now Facebook Wants My Credit Card Number. Available at: http://aniruhama.wordpress.com/2006/09/13/now-facebook-wants-my-credit-card-number/ [Accessed August 16, 2008].

Schonfeld, E., 2007. Social Site Rankings (September, 2007). TechCrunch. Available at: http://www.techcrunch.com/2007/10/24/social-site-rankings-september-2007/ [Accessed July 26, 2008].

Sheth, A. et al., 2005. Semantic Association Identification and Knowledge Discovery for National Security Applications. Journal of Database Management, 16(1), 33-53.

Siegel, L., 2008. *Against the Machine: Being Human in the Age of the Electronic Mob*, Spiegel & Grau.

Simon, H.A., 1977. The New Science of Management Decision, Prentice Hall PTR Upper Saddle River, NJ, USA.

Simon, J., 2006. Chase Credit cards will be profiled on Facebook.com. Available at: http://www.creditcards.com/credit-card-news/chase-credit-cards-to-appear-on-facebook.php [Accessed August 16, 2008].

Shiels, M., 2008. Facebook 'violates privacy laws'. BBC. Available at: http://news.bbc.co.uk/2/hi/technology/7428833.stm [Accessed June 2, 2008].

Smith, R.D., 2002. Instant Messaging as a Scale-Free Network. Arxiv preprint cond-mat/0206378.

Soghoian, C., 2008a. Exclusive: The next Facebook privacy scandal. Available at: http://www.cnet.com/8301-13739_1-9854409-46.html.

Soghoian, C., 2008b. Interview with Chris Soghoian about article 'Exclusie: The Next Facebook Privacy Scandal'.

Solove, D.J., 2006. A Taxonomy of Privacy. *University of Pennsylvania Law Review*, 154(3), 477-560.

Solove, D.J., 2007d. Anonymity and Cyber-Bullies. Concurring Opinions. Available at: http://www.concurringopinions.com/archives/2007/11/anonymity_and_c.html [Accessed August 22, 2008].

Solove, D.J., 2007c. Facebook and the Appropriation of Name or Likeness Tort. Concurring Opinions. Available at: http://www.concurringopinions.com/archives/2007/11/facebook_and_th.html [Accessed August 7, 2008].

Solove, D.J., 2007. Facebook's Beacon, Blockbuster, and the Video Privacy Protection Act. Available at: http://www.concurringopinions.com/archives/2007/12/facebooks_beaco_1.html#comments [Accessed April 17, 2008].

Solove, D.J. & Rotenberg, M., 2003. *Information Privacy Law*, Aspen Publishers.

Solove, D.J., 2008a. Interview about article 'A Privacy Taxonomy' and books 'The Digital Person' and 'The Future of Reputation'.

Solove, D.J., 2007e. Megan Meier Blog: A Hoax. Concurring Opinions. Available at: http://www.concurringopinions.com/archives/2007/12/megan_meier_blo.html [Accessed August 22, 2008]

Solove, D.J., 2008c. Megan Meier Case Update -- Drew Indicted. Concurring Opinions. Available at: http://www.concurringopinions.com/archives/2008/05/megan_meier_cas.html [Accessed August 22, 2008].

Solove, D.J., 2007f. More Facts about the Megan Meier Case. Concurring Opinions. Available at: http://www.concurringopinions.com/archives/2007/12/more_facts_abou.html [Accessed August 22, 2008].

Solove, D.J., 2008d. More Misguided Responses to the Megan Meier Incident. Concurring Opinions. Available at: http://www.concurringopinions.com/archives/2008/05/more_misguided.html [Accessed August 22, 2008].

Solove, D.J., 2008. Privacy, Free Speech and Anonymity on the Internet Part 2 of Discussion. *Washington Post*, Internet.

Solove, D.J., 2007f. Should Megan Meier's Tormentors Be Shamed Online? Concurring Opinions. Available at: http://www.concurringopinions.com/archives/2007/11/should_megan_me.html [Accessed August 22, 2008].

Solove, D.J., 2004. *The Digital Person: Technology and Privacy in the Information Age*, NYU Press.

Solove, D.J., 2007a. *The Future of Reputation: Gossip, Rumor, and Privacy on the Internet*, Yale University Press.

Solove, D.J., 2007g. The Megan Meier Case: New Developments. Concurring Opinions. Available at: http://www.concurringopinions.com/archives/2007/12/megan_meier_and.html [Accessed August 22, 2008].

Solove, D.J., 2007b. The New Facebook Ads -- Starring You: Another Privacy Debacle? Concurring Opinions. Available at: http://www.concurringopinions.com/archives/2007/11/the_new_faceboo.html [Accessed August 7, 2008].

Sophos, 2007. Sophos Facebook ID probe shows 41% of users happy to reveal all to potential identity thieves. Available at: http://www.sophos.com/pressoffice/news/articles/2007/08/facebook.html [Accessed August 22, 2008].

Sophos, Facebook - Sophos investigates the privacy and productivity challenge. Available at: http://www.sophos.com/security/topic/facebook.html [Accessed August 17, 2008].

Stelter, B., 2008. The Facebooker Who Friended Obama. The New York Times. Available at: http://www.nytimes.com/2008/07/07/technology/07hughes.html?_r=1&scp=1&sq=facebook%20obama&st=cse&oref=slogin [Accessed July 27, 2008].

Strahilevitz, L.J., 2005. *A Social Networks Theory of Privacy*, Law and Economics Programme, Faculty of Law, University of Toronto.

Strogatz, S.H., 2001. Exploring complex networks. Fo, 19, 53.

Stutzman, F., 2006. An evaluation of identity-sharing behavior in social network communities. *Online proceedings of the 2006 iDMAa and IMS code conference*.

Sunden, J., 2003. Material virtualities: approaching online textual embodiment, Peter Lang Publishing.

TechNews, 2006. Social Networking Exposes You To Hackers and Identity Thieves. Available at: http://www.technologynewsdaily.com/node/4696.

Tien, L., 2008. Privacy Harms for Users of Social Network SItes.

The Reporters Committee for Freedom of the Press, 2003. Handbook: Invasion of Privacy: Publication of private facts. Available at: http://www.rcfp.org/handbook/c02p03.html [Accessed August 17, 2008].

US Patent & Trademark Office, 2007. United States Patent Application: 0070192299. Available at: http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnetahtml%2FPTO%2Fsearch-adv.html&r=1&p=1&f=G&l=50&d=PG01&S1=20070192299.PGNR.&OS=dn/20070192299&RS=DN/20070192299 [Accessed April 8, 2008].

Vallance, C., 2008. Facebook faces privacy questions. BBC. Available at: http://news.bbc.co.uk/2/hi/technology/7196803.stm [Accessed April 9, 2008].

Verdi, J., 2008. Interview with John Verdi about his work on 'Third party applications for Facebook' and Social Network Sites at EPIC.

Verkalk, R., 2008. Mother sues over tale of 'drunken party' lifted from Bebo. Available at: http://www.independent.co.uk/news/uk/home-news/mother-sues-over-tale-of-drunken-party-lifted-from-bebo-865039.html [Accessed August 17, 2008].

Walzer, M., 1983. Spheres of Justice: A Defense of Pluralism and Equality, Basic Books.

We Feel Fine, We Feel Fine / Methodology. Available at: http://wefeelfine.org/methodology.html [Accessed August 24, 2008].

Wikipedia, 2008a. Principle of least privilege —- Wikipedia, The Free Encyclopedia.

Wikipedia, 2008b. Security through obscurity —- Wikipedia, The Free Encyclopedia.

Zittrain, J., 2008. *The Future of the Internet--And How to Stop It*, Yale University Press.

Zittrain, J., 2000. What the Publisher Can Teach the Patient: Intellectual Property and Privacy in an Era of Trusted Privication. *Stanford Law Review*, 52(5), 1201-1250.

Zuckerberg, M., 2007. Thoughts on Beacon. Available at: http://blog.facebook.com/blog.php?post=7584397130 [Accessed April 8, 2008].

III. Appendices

- Appendix A – Facebook Case Study
- Appendix B – Interviews
  - Interview Daniel Solove
  - Interview Chris Soghoian
  - Interview Adrienne Felt
  - Interview Guilherme Roschke
  - Interview Jeroen van den Hoven
  - Interview John Verdi
  - Interview Lee Tien
- Appendix C – Minutes Graduation Committee
  - Minutes Kick-off Meeting
  - Minutes Midterm Meeting
- Appendix D – Set up of survey
- Appendix E - Survey
- Appendix F – All results
- Appendix G – Analysis of data
- Appendix H – List of prioritized privacy harms, aggregated.
- Appendix I – List of all privacy harms, aggregated.
- Appendix J – Model for calculating gamma.
- Appendix K – Print of all descriptive statistics for the tort laws
- Appendix L – Print of all descriptive statistics for the privacy harms
- Appendix M – Table of rater agreement for tort laws
- Appendix N – Table of rater agreement for probability and impact
- Appendix O - Presentation and Workshop at CFP 2008
- Appendix P - Weblogs
- Appendix Q - Actor Analysis