D4.2

User Categorisation of Digital Footprint – v1

V1.4 / 2014-10-27

Tom Seymoens (iMinds), Laurence Claeys (iMinds), Jo Pierson (iMinds), Sanne Ruelens (iMinds), Jo Pierson (iMinds), Adrian Popescu (CEA)

This deliverable discloses the methodology behind the upcoming quantitative and qualitative research track that will result in a user categorisation of their digital footprint. As such it will describe the technical solutions provided for both research tracks as well as how they are related to the different work packages.

Project acronym	USEMP
Full title	User Empowerment for Enhanced Online Presence Management
Grant agreement number	611596
Funding scheme	Specific Targeted Research Project (STREP)
Work program topic	Objective ICT-2013.1.7 Future Internet Research Experimentation
Project start date	2013-10-01
Project Duration	36 months

Workpackage	WP4
Deliverable lead org.	IMinds
Deliverable type	Report
Authors	Tom Seymoens (iMinds) Laurence Claeys (iMinds) Jo Pierson (iMinds) Adrian Popescu (CEA)
Reviewers	Symeon Papadopoulos (CERTH) Noel Catterall (HWC)
Version	1.4
Status	**Final**
Dissemination level	**PU: Public**
Due date	2014-09-30
Delivery date	2014-10-27

Version	**Changes**

V1.1	TOC + Abstract
V1.2	Literature study + Qualitative Research
V1.3	1st Full Version
V1.4	Final Version with input Reviewers

Table of Contents

1.Introduction

In Task 4.2 of WP4, we aim to enhance our understanding of how users value their personal data on Online Social Networks. On these websites, users can disclose demographic information, update their status, share emotions and thoughts, post photos and videos, and share personal interests (Feng & Xie, 2014). Besides volunteered data, a second category of personal data can be differentiated on online platforms, namely: observed data. This type is created as a result of a transaction between an individual and an organization (e.g. Location data from a mobile phone, credit card transactions, purchase history at a retailer, clickstreams on a website etc.) (World Economic Forum, 2012). These activities leave a footprint that is searchable and traceable by advertisers and other third parties (Feng & Xie, 2014). Through the analyses and combination of both types of data, one can derive more in-depth predictions of preferences of the user, such as purchase intent, health or even financial data. This is called inferred data (World Economic Forum, 2012).

In our previous research it was exposed that although there was general awareness towards volunteered data among the end-users, awareness of observed data and most certainly inferred data was considerably lower (see USEMP deliverable 4.1: Social Requirement Analysis – v1). We saw that as the data gets more revealing, and as such possibly more sensitive, people became less aware of its existence. In our aim of making the economic logic and processes behind social platforms more transparent, we first need to fully understand what information users consider 'sensitive' or 'private' information and their sentiment towards its disclosure.

This brings us to the scope of this deliverable. In what follows, we describe the different methods we use to come up with a user-defined categorisation of private OSN data. First, we examine what type of information users generally upload to Facebook and whether or not this leads to conclusions on what users consider to be private and public information. We present the work done in a literature study of previous research on information disclosure on Facebook. Next, we elaborate on our upcoming quantitative research where we examine private profiles of users who agreed upon opening up their data for USEMP. Here we aim to see if it is possible to infer sensitive information from the data users provide on Facebook. Specifically, we will see if we can derive insights on the eight privacy dimensions derived from the work done in WP6: demographics, psychological traits, sexual preferences, political attitudes, religious beliefs & cultural heritage, health factors & condition, location and consumer profile (The privacy dimensions will be discussed in depth in D6.1, due date month 15). A questionnaire will be used to verify if our predictions are correct and to train our inference algorithms.

Furthermore, we outline a qualitative research track where we hope to identify how users value their privacy and the disclosure of personal data. We will do so by setting up interviews, making use of the Q-card sorting method.

In the last part of this deliverable we take a look at how the industry values personal data. This relates to T3.7, where we try to place USEMP in the existing value network and how it will change under influence of the upcoming EU regulation.

2. Literature Study: Information Disclosure on Facebook

2.1. Introduction

In this chapter of the deliverable, we will present the work done of a literature study exploring the following questions:

- Which data do users consciously reveal on their social network platforms?
- What are the users' main motivations for disclosing this information?
- What are the users' main motivations for retaining information?
- Is there an evolution on the overall degree of information-sharing practices?

We first take a look at general user practices, after which we try to answer our main questions one by one for the demographics of teenagers and adolescents. This distinction came forth from the literature, where university students were the most studied. Much less attention was given towards adults older then 25. We should take this into consideration when conducting our own research.

Because the (privacy) features and configuration parameters in social media are constantly evolving, the studies are to be interpreted within the timeframe they were executed. We therefore acknowledge that some of the data in this literature overview can be out-dated, as some of it dates back to 2009. Nevertheless, it will be interesting to compare these findings with the results of our upcoming quantitative research.

2.2. General user practices concerning information disclosure

2.2.1. Personal Information Disclosure

In their survey about the Facebook usage of 1000 randomly selected university community members (including staff, students and faculty), Hoadley, Xu, Lee and Rosson (2010) found that most respondents are selective in terms of the type of personal information they post on Facebook. The majority post **photos** and reveal their **sexual orientation**, **relationship**

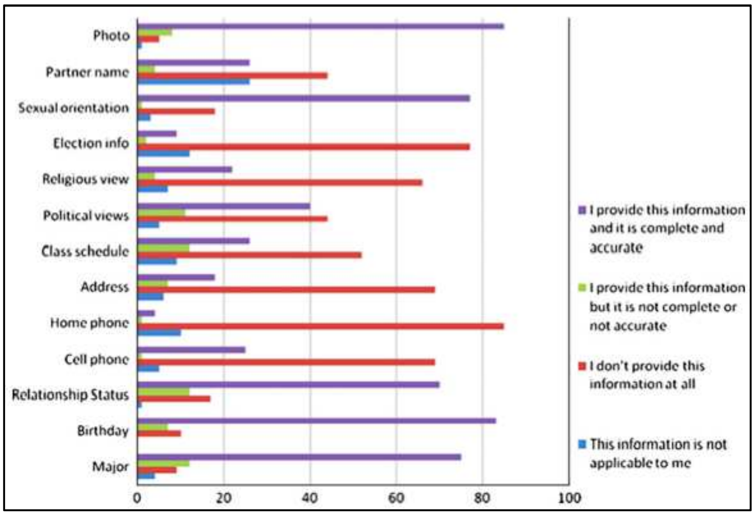


Figure 1: Percentages of respondents' profiles revealing personal information

status and **birthday**. A remark that could be made concerning sexual orientation is that people who do not share it might have a greater chance to be not-heterosexual, as a minority might be less likely to disclose their preference. This could be interesting to research. Many users did conceal their **political affiliations**, **religious views**, **address** and **phone numbers**. (See Figure 1 (Hoadley et al., 2010)).

These findings correspond with the research done by Nosko, Wood and Molema (2010), where 400 randomly selected, accessible, personal profiles were collected to study which types of information were most likely to be disclosed on social network sites. The participants' ages ranged from 19 to 47 years old.

In their study they found that the **15 most consistently disclosed pieces of information** (on more than 63% of the profiles) could be put under the following categories:

- Personally identifying information: e.g. birth date, gender, profile pictures
- Social connections: e.g. which groups were joined, which friends they had
- Education information
- Regular update information (status, wall posts, etc.)

The 15 least provided pieces of information (available on less than 9% of the profiles) described:

- Key personal information (zip codes, home address, phone number, etc.)

2.2.2. Reasons for information disclosure

Nosko et al. (2010) saw two interesting trends in their results. First, they noticed that as age increases, the user releases lower levels of personal information. An explanation for this might be that older generations have less trust in the technologies at hand or that they grew more wary towards the disclosure of private information.

Secondly, they detected that individuals who were single share the greatest amount of what they call stigmatizing information: religious and political views, sexual orientation, interests and media preferences. Nosko et al. (2010) explain this as follows: "Those seeking a relationship may be using Facebook as a less overt dating site, and, thus may be differentially motivated to disclose highly personal information across a variety of topics regardless of the dangers or threats associated with disclosing this information". These motivations for information disclosure need to be further investigated.

2.2.3. Reasons for information retention

Dwyer, Hiltz, & Passerini (2007) investigated the impact of trust and Internet privacy concern on information sharing. They used a survey, which they tried to disseminate through public groups on Facebook and MySpace. Eventually 117 subjects participated in their research, of which 69 were members of Facebook and 48 used MySpace. Where they found little evidence for a correlation between privacy concern and information disclosure, they did find that trust in Facebook correlates with the sharing of cell phone numbers. Further research on how trust affects information disclosure is needed.

In a study by Das, & Kramer, (2013) they found an underlying principle for self-censorship on Facebook. People seem to censor themselves more when they are not completely aware of the extent of their audience. They then tend to make sure that their posts are appropriate for the lowest common denominator.

Two other reasons for not disclosing were given by Bevan, Gomez, & Sparks (2014). The first motivation was self-protection, measured as in a survey as e.g.: “I decided not to share this news about an important event with my Facebook friends because I might get hurt”. Secondly friend unresponsiveness was brought to the forefront as another reason to not reveal certain information.

2.2.4. Evolution in information disclosure

For their research, Stutzman, Gross, & Acquisti (2013) collected a longitudinal database of publicly disclosed information on Facebook profiles. Their dataset started with just over 3000 profiles and grew to over 20.000. They noticed how Facebook users since 2005 steadily kept reducing the amount of information they disclose publicly with strangers (with an exception in 2009 and 2010, assumingly due to changes in Facebook’s policy and interface). However, in the same period they did reveal more and more information with their connected profiles (friends).

Stutzman et al. (2013) propose a couple of interesting reasons for this evolution:

- Since 2007, new data can be generated by third-party applications (e.g. songs played on Spotify)
- The occurrence of data that other users reveal about you: e.g. by tagging you in pictures and locations
- Through the availability to target certain user groups, people gained the feeling of more control and felt safer sharing online.
- Profiles have been changed from “[…] a static representation of personal information (gender, name) to habitats through which new information is frequently created (places visited, events attended)” (Stutzman et al., 2013, p. 20)

Stutzman et al. (2013) notice that, “[…] the increased private disclosures ended up reaching entities other than a user’s friends”, such as Facebook, third-party apps, Facebook advertisers, often without awareness or explicit content.

2.2.5. Dimensionality of information disclosure

To conclude this part on the general user practices, we like to mention an interesting observation made by Knijnenburg, Kobsa, & Jin (2013). They noticed how people are often classified into three main segments based on their overall degree of information disclosure: privacy fundamentalists, pragmatists and unconcerned. They propose that information disclosure behaviour is not one-dimensional, but that people can be differentiated by which kind of information they tend to reveal (Knijnenburg, 2014). For example, they were able to differentiate between individuals that have:

- Low intention to disclose contact information and Facebook activity, but high intention to disclose their location and interest.

And individuals that have:

- Low intention to disclose contact information and location, but high intention to disclose Facebook activity and interests.

This is an important observation when developing privacy enhancing tools, because this means that “[…] groups of people with the same amount of overall disclosure can show very different “disclosure profiles” […](Knijnenburg et al., 2013, p. 29). As a consequence, privacy enhancing tools need to be adapted to these different user needs in order to be successful.

2.3. Teenagers and Young Adults’ Information Disclosure

In the coming paragraphs the results of several studies towards teenagers and adolescents’ information disclosure are set next to each other. We defined teenagers as those aged 12 to 17, young adults were defined in the age range: 17 to 25.

A study published by the PEW Research Center in 2013 examines teens’ privacy management strategies on social media sites (Madden et al., 2013). Their findings are based on a survey of 802 teens, aged 12 to 17, which was conducted from July 26 until September 30, 2012. The results presented about teens’ information disclosure are primarily based on this study, where this is not the case this will be explicitly stated.

2.3.1. Public versus private profiles

In the PEW study, most teenagers (60%) keep their social network profiles private. This means that they only grant access to the people they friended. 25% also allow access to friends of friends, and only 14% keep their profiles completely public. Interestingly enough, gender differences could be discerned (70% of the girls keep their profiles completely private compared to 50% of the boys). Also, only 8% of the girls claimed to have a completely public profile, whereas this is the case for 20% of the boys.

Taraszow, Aristodemou, Shitta, Laouris, & Arsoy (2010) found that in the age group of 18 to 22 the majority of the respondents (64,1%) had a private Facebook profile.

2.3.2. Personal Information Disclosure

Let’s take a look at which types of information are often available on teenagers’ (aged 12 to 17) social network profiles (Madden et al., 2013, p. 30).

- 92% use their **real name**
- 91% post a **photo of themselves**
- 84% post their **media preferences** (such as movies, music or books they like)
- 82% post their **birth date**
- 71% post their **school name**
- 71% post the **city or town where they live**
- 62% post their **relationship status**
- 53% post their **email address**
- 24% post **videos of themselves**
- 20% post their **cell phone number**
- 16% have their profile set up to **automatically include location** in their posts.

When we compare these results to a study by Young & Quan-Haase (2009), where 77 undergraduate university students were asked to fill in a survey on their information disclosure on Facebook. Their ages ranged from 17 to 25.

- 99,35% of the participants used their **real name** on Facebook.
- 98,7% provide **images of themselves**
- About 66% revealed their **sexual orientation**

For sake of completeness, we'll add the results of a study of 343 undergraduate students in Ontario, Canada (Christofides, Muise, & Desmarais, 2009):

- 96% post their **birth date**
- 86% post their **email address**
- 85% post their **hometown**
- 81% post their **relationship status**
- 72% post their **school name and programme**
- 24% post their **phone number**
- 4% post their **home address**
- The participants were also very likely to post a **profile picture and pictures with friends**

In Table 1, we compare the disclosure behaviour of the seven types of information we found information about for both teenagers and young adults.

	Teenagers	**Young Adults**
Real name	92%	99,35%
Pictures of themselves	91%	98,7%
Birthdate	82%	96%
School name	71%	72%
Phone number	20%	24%
Relationship Status	62%	81%
Email address	53%	86%

Table 1: Comparing Teens and young adults' Information Disclosure

2.3.3. Evolution in Information Disclosure

Teenagers were also questioned about the disclosure of five of the above information types (photo of themselves, school name, city or town they live in, email address and cell phone number) in a PEW study performed in 2006. A comparison between the results brought to light that for each of the five types of information the disclosure became more common, as is illustrated by Figure 2 (Madden et al., 2013). They provide three possible reasons for this increase in information sharing:

- A change of dominant social networking platform. In 2006, MySpace was still widely used and it had a different form than the current most frequently used platform, Facebook.
- The arrival of new devices that are often used for sharing personal information (smartphones, tablets, etc.).
- Interface changes on the social platforms that encourage users to disclose more information.

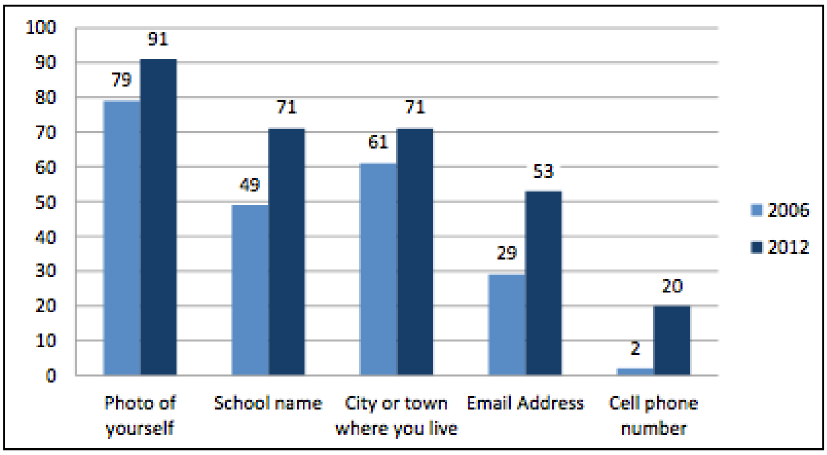


Figure 2: Social media profiles: What teens post - 2006 vs 2012 (in %)

2.3.4. Reasons for information disclosure

Teenagers link Facebook with social interaction and self-expression and not with information sharing or privacy. As such, they see the amount of likes they get and friends they have as a proxy for social status (Madden et al., 2013). This might explain that their main concern is not keeping their audience to a minimum.

Moreover, personal information such as birth date and city or town they live in is often required when signing up. Facebook even forbids the practice of fake names in its Terms of Service.

When we look at the slightly older age group of **young adults**, we see that their main motivation for self-disclosure corresponds with what drove the teens:

Pempek et al. (2009) observe how media use can address some of the challenges that adolescents encounter when developing their identity. One way this happens is by providing means for self-disclosure. Therefore, this might explain why adolescents share their religious beliefs, political ideology, and work – which are the classic identity markers – to present and define themselves toward others. In the same study by Pempek et al. (2009), media preferences came forth as a new category of personal information that individuals want to share to express themselves.

The results of a study by Christofides et al. (2009) agree with these findings as they observed that the need for popularity is a significant predictor of information disclosure among adolescents (Christofides et al., 2009). They explained this as follows: “The risks of limiting access to personal information become greater than the risks of disclosure, because when limiting access, the individual also limits the potential for identity construction and thus potentially reduces his or her popularity” (Christofides et al., 2009, p. 343).

2.3.5. Reasons for information retention

First we’ll take a short look at the reasons given by the PEW study why **teenagers** would retain certain information from their profiles. Some general motivations were listed, after which they investigated if privacy concern and awareness towards third party access might influence their disclosure behaviour.

A) General reasons for information retention

- Parental influence
- A previous bad (online) social experience which made them change their privacy settings

B) Concerns and awareness about third party access

- Third party access is generally not one of the reasons why teenagers might retain information from the online sphere. According to Madden et al. (2013), some teens do not know if other parties use their information: “Without […] seeing what those negative experiences might be, teens do not seem to be overly concerned about advertisers and third parties having access to their information.”

For **young adults**, the main reasons for disclosing information (self-expression, identity construction and popularity) are inherently connected to the main reason for retaining information. As the study of Christofides et al. (2009) points out that individuals with higher self-esteem more frequently use information control tools (e.g. that can limit audiences) and thus do not share information with everyone as “[…]they are only concerned about their popularity within their chosen circle.”(Christofides et al., 2009, p. 343). In this sense it becomes apparent that groups with lower self-esteem might more quickly make use of online social network sites to promote their identity to larger audiences.

3.Quantitative research track

The previous chapter provided an overview of scientific research done in the past on the information disclosure practices of users on social network sites, mainly Facebook. We got some insights about which personal information users generally disclose on their profiles, the reasons for these revelations and how this has transformed over time. Two remarks have to be made in consideration of this literature study. First of all, the social media environment is a rapidly changing one. A good example of this is how the amount of monthly active Facebook users doubled between 2010 and 2013, from 608 million to 1,230 billion. Another example is the increasing awareness of users about privacy issues with data shared on social networks and, more generally, on the Web. Therefore, although it brings valuable insights, previous scientific research has to be updated and completed in order to take social network dynamics into account. A second remark is that most of existing literature deals with the disclosure of data that users consciously submitted themselves, i.e. volunteered data. In USEMP, we want to maintain a broader perspective and, in the upcoming quantitative research track, we analyse both the volunteered and observed OSN data of end-users, to see if we can infer more in-depth information, of which the user may not be aware. Our initial idea was to do this in two different ways.

A) Quantitative analyses of private Facebook profiles

In this research track, we will ask Facebook users to open up their profiles. In this way we can get access to their volunteered and observed data, such as location information, and explore what other information can be inferred from their disclosure. The participants are also asked to fill in a survey which focuses on eight privacy dimensions: demographics, psychological traits, sexual preferences, political attitudes, religious beliefs & cultural heritage, health factors & condition, location and consumer profile.

This research track brings added value in three complementary directions. First, we get an up to date overview of what type of information users typically disclose on their Facebook profiles. We can compare these results with the results reported in existing literature. Second, based on the data we try to infer some personal information of sensitive nature. We do this by applying the multimedia information extraction and personal information mining algorithms created in WP5 and WP6 of the project. The accompanying survey is key for these algorithms, since it provides the data necessary for training and validating a part of the algorithms. Third, through means of the survey we learn which items our participants find very sensitive and if they think this information can be found on their social network profile. We can later combine this information with the results from the algorithmic processing. By doing this we can make the user more aware of the degree of their information disclosure and the ability of third parties to access this info.

In the 3.1.1 – 3.1.4., we will provide a more detailed description of the practical set-up of this research track. The recruitment strategy, technical solution, accompanying questionnaire and timeline will be presented here.

B) Quantitative analyses of public available Facebook profiles

The second approach we initially proposed in the description of work was a large-scale statistical analysis of publicly available Facebook profiles. This analysis was meant to be fully anonymous and, since the profiles were publicly available from Web search engines, did not require explicit consent from users. The idea was to get access and examine a substantial set of user profile data to underwrite the findings of our smaller-scale analyses and qualitative Q card interviews. This would have helped our understanding of privacy practices on the Web and defining a user categorisation of their digital footprint. Eventually, concerns related mainly to ethical, legal and technical aspects, made us reconsider this process.

Although it would advance our understanding of privacy issues, a key objective of the project is to give the users more control over their personal data. From an ethical point of view, it is thus delicate to make use of this public data without the prior consent of their creators. Furthermore, we want to make the end-user **more aware** towards the drawing of inferences by third parties, doing this ourselves without their approval seems to conflict with this principle.

From a legal perspective, Facebook's "Statement of Rights and Responsibilities" forbids third parties to "collect users' content or information, or otherwise access Facebook, using automated means (such as harvesting bots, robots, spiders, or scrapers)" without Facebook's prior permission[1] (Section 3.2). While technically possible to some extent, automatic access to a large set of public user profiles would go against Facebook's policies and could lead to an important legal vulnerability of the project.

Considering the issues cited above, the small-scale quantitative and qualitative research are the more advisable ways to go. Here the end-users are informed of the data gathering process, which gives them more control as they can choose to open up their data to USEMP tools. These research tracks also give us the ability to inquire how users feel about the fact that inferences are being drawn about their consumer profile, sexual orientation, etc. and make them more aware of how this happens.

This is also the reason we want to provide the user clear and simple information in a Data License Agreement about what they commit to, before they participate in our research:

"(G) You hereby grant Your consent to process Your sensitive personal data, notably those revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and those concerning health or sex life."

The work done in a combination of the small-scale quantitative data analyses and qualitative Q card method presents us with an alternative to get a view on how users categorize and value their personal data. In addition, USEMP tools will be open to the general public in the final year of the project. If a sufficiently high number of users subscribe to the service, the scale of quantitative data analysis could significantly increase at that time and, in this eventuality, we will update the associated study with new insights gained from this larger pool of users.

[1] https://www.facebook.com/legal/terms

3.1. Research Plan – Quantitative Research Track

In this part, a comprehensive overview is presented of the strategy used for implementing the upcoming quantitative research track. With this research, we aim to build an understanding of which user data can generally be found on online social networks and to refine the developed multimedia information extraction and private information mining algorithms that were are currently engineered as part of WP5 and WP6. We will make use of the DataBait Research tool, which will be presented in the following paragraph, based on wireframes.

3.1.1. Setting up the technical solution – The DataBait Research tool

USEMP tools will be implemented in a Web application, named Databait. This application will notably enable the following interactions:

- Subscribing with an existing Facebook account,
- Filling in a questionnaire about core privacy dimensions (compulsory for pilot participants, optional for other subscribers),
- Visualizing the privacy status along different dimensions,
- Visualizing a consumer profile and monetisable data,
- Getting details about particular items which have a strong influence on privacy.

User data are gathered through the following channels:

- Questionnaire – to get explicit information about different privacy dimensions, which is exploited for training data mining algorithms (only for a subset of users),
- Facebook API plug-in – to gather data which was voluntarily shared by the user. The USEMP app needs to be approved by Facebook and, in case this process is problematic, a fall-back solution consists in asking users to download and provide their Facebook historical data.
- Browser plug-in – to get behavioural data about the user's interaction with OSNs and other websites.

The user interactions and the data gathering process are detailed in deliverable D7.2, due at month 12.

One central objective of USEMP is to infer higher-level knowledge from the data shared by the users. To attain this objective, two main types of algorithms are implemented. In WP5, the project develops multimedia (text and images) information extraction tools in order to process information, which is shared voluntarily. In WP6, the focus is on privacy estimation algorithms, which exploit both voluntary and observed user data.

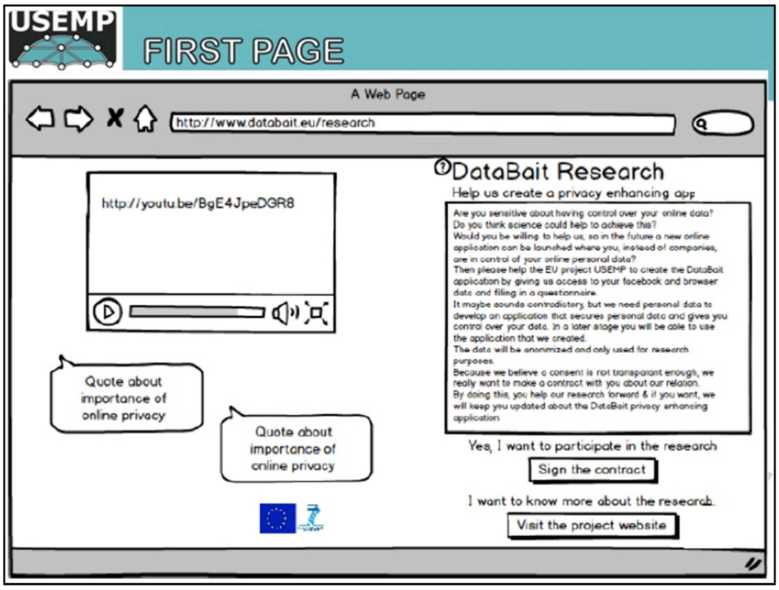
Given the strongly sensitive character of the data, focus will be put on the security of their transmission and storage. To strengthen security, all Databait components and users' data will reside in HWC premises, a partner with strong expertise in handling personal data. Some of the data will be also transferred to CERTH and VELTI in order to train privacy prediction algorithms and to perform visualisation, respectively. These transfers will be performed on a per need basis and the local storage will be limited to the period necessary for processing related to USEMP purposes. Similar to HWC, state of the art security standards will be implemented by both CERTH and VELTI in order to ensure data security.

3.1.2. Presenting the DataBait Research tool

The users in our research will have to go through the following different steps in order for their participation to be useful for us.

1) The participant visits the website of the Research tool

Here s/he gets presented with some more information on the goal of the research and why their participation is of greatest value to us. The user gets referred to the website of the USEMP-project, if they are not convinced yet and require more information.


USEMP
FIRST PAGE
A Web Page
http://www.databait.eu/research
http://youtu.be/BgE4JpeDGR8
Quote about importance of online privacy
Quote about importance of online privacy
DataBait Research
Help us create a privacy enhancing app
Yes, I want to participate in the research
Sign the contract
I want to know more about the research
Visit the project website

Are you sensitive about having control over your online data?
Do you think science could help to achieve this?
Would you be willing to help us, so in the future a new online application can be launched where you, instead of companies, are in control of your online personal data?
Then please help the EU project USEMP to create the DataBait application by giving us access to your facebook and browser data and filling in a questionnaire.
It maybe sounds contradictory, but we need personal data to develop an application that secures personal data and gives you control over your data. In a later stage you will be able to use the application that we created.
The data will be anonimized and only used for research purposes.
Because we believe a consent is not transparant enough, we really want to make a contract with you about our relation.
By doing this, you help our research forward & if you want, we will keep you updated about the DataBait privacy enhancing application

Figure 3: Wireframe of the DataBait Research tool

2) The user gets presented with the Data License Agreement

If the user is interested in partaking, he continues to the next page where he gets presented with the Data License Agreement, which was created to simply state why we need user data to build a tool to help users manage their personal data on the Web. Here they can agree to allow us to gather their Facebook data.

3) The user Creates a DataBait account & logs in to his/her Facebook account

With logging in to their Facebook account the user grants us the access we need for gathering their Facebook data. This is a crucial step in the process.

4) The user fills in the questionnaire

This step will not be compulsory, as the sensitive nature of the questions might scare some users that wanted to open up their Facebook data. The survey gauges information that is related to the eight privacy dimensions. The answers are used to ameliorate the inference algorithms. More information on the survey can be found under 3.1.4.

5) User is thanked for involvement

The last screen the participants receive is an important one. They get the option to keep up to date on the launch of the DataBait tool. This makes it possible to keep them involved throughout the different versions of the tool and may provide useful feedback for updates.

3.1.3. User Recruitment

The biggest goal of the DataBait Research tool is to give the USEMP researchers data they can work upon to refine the developed inference algorithms. The short-term value for the users to install this tool is small. The users will most probably participate in this part of the research due to their willingness to bring research forward and their belief that by helping USEMP in this stage of the project, a new interesting PET will be created. It is also in this manner that we will approach potential participants.

Our aim is to have access to the Facebook data of 300 users with the following criteria:

- Active Facebook users
- Computer & Internet access at home
- Willing to share Facebook data for research purposes

We will make use of the two living lab organisations involved in the project (Belgian's iLab.o[2] and Sweden's Botnia[3]) to each deliver 50 users for participation in this quantitative research track. These users are expected to stay involved in the USEMP project until the end, so we will also make use of them in both field trials. To keep them committed throughout the project they will get regular updates and information about the evolution of the project.

The 200 other participants will be gathered by making a social media campaign in Belgium and Sweden, involving friends, relatives, university students, staff, etc.

3.1.4. Questionnaire

A) Privacy Dimensions

The questionnaire that accompanies the DataBait Research tool is based on the privacy dimensions that were identified as part of WP 6. "Privacy dimensions reflect user traits and information that are typically considered sensitive, but are often extensively used by third parties for profiling and targeting or even for decision making that should normally not rely on such data e.g. for declining insurance, rejecting candidates etc." (Internal document privacy dimensions – WP6). Eight privacy dimensions were defined, each with a specific set of variables that can possibly be used to characterize. The main purpose of the questionnaire at hand is to validate the inferences made by the algorithms that predict the variables leading to the following privacy dimensions: Demographics, Psychological Traits, Sexual Profile, Political Attitudes, Religious Beliefs, Health Factors and Condition, Location and Consumer Profile.

A) Questionnaire

Presented in our annex is the accompanying questionnaire as it is currently made up. An ethics commission of one of our partners is reviewing this, since it asks some sensitive information. After every question, the respondent can also stipulate whether they find this sensitive information or not and whether they think this information is available on their social network profile. We'll shortly provide here the different variables that were defined as cues

[2] http://www.openlivinglabs.eu/livinglab/iminds-ilabo
[3] http://www.openlivinglabs.eu/node/125

for characterizing the eight privacy dimensions. For more information, on how these came about we refer to D6.1

1) **Variables for demographics**
    a. Age
    b. Gender (male/female)
    c. Ethnicity (or nationality)
    d. Literacy level
    e. Occupation
    f. Income Level
    g. Family status (married/single, number of children)

2) **Variables for psychological traits**
    a. Emotional stability
    b. Agreeableness
    c. Extraversion
    d. Conscientiousness
    e. Openness

3) **Variables for sexual Profile**
    a. Relationship status (in a relationship/single)
    b. Sexual preference (heterosexual/homosexual/bisexual)
    c. Multiple partners
    d. Habits

4) **Variables for political attitude**
    a. Political parties
    b. Politicians
    c. Stance in issues
5) **Variables for religious beliefs and cultural heritage**
    a. Supported religion

6) **Variables for health factors and condition**
    a. Smoking behaviour
    b. Drinking behaviour
    c. Drug use
    d. Chronic diseases
    e. Mediating factors (e.g. exercising)
    f. Medical history

7) **Variables for location**
    a. Home address
    b. Work address

c. Favourite places
d. Visited places

8) **Variables for consumer profile**
   a. Preferred products
   b. Brand attitude
   c. Hobbies
   d. Devices

4. Qualitative research track

Next to our quantitative exploration of user data, explained in the previous chapter, we will now elaborate on how we will make use of qualitative interview methods to get a view on the valuation of personal data and privacy, not only as seen by the user, but from the industry perspective as well.

4.1. Research plan – User Perspective

As part of previous research in the USEMP-project, we conducted four focus groups (See deliverable 4.1). Here we noticed that the participants understood to a certain degree that the data they inserted online was being collected. They primarily mentioned the big new media players, such as Facebook and Google, as the organisations for which their personal data have an economic value. This was an accurate observation, as the selling of user data is the business model that underlies the apparent free provision of a variety of services. Carrascal, Riederer, Erramilli, Cherubini, & de Oliveira (2013) look at the market for personal data as a two-sided platform market, where you **have** a business serving two or more distinct types of customers who depend on each other in some important way, and whose joint participation makes the platform valuable to each (Rochet & Tirole, 2004). On the one hand the marketing companies and on the other hand the user, and both feed the (social media) platform. They notice that in such a system it is possible for the service-provider/platform to attach value on personal information, based on the revenues they get from selling the data, but for the user it is more difficult to do this. They distinguish between two types of difficulties when validating data by users:

1. **Based on context:**
   - **a.** Type of information: e.g. users might perceive the data from a health-related online search more valuable than this of the search for a new pair of boots.
   - **b.** Type of interaction: e.g. financial interactions might be identified as more valuable than social interactions.

2. **Based on personal demographics:**
   - **a.** The user's education level, socio-economic status, age, gender, … might have an influence on how they value their data.

This consciousness about the value of personal data is beneficial for the user's cost-benefit analyses between loss of privacy and the service they get in return. Also, Acquisti, John, & Loewenstein (2013) see how businesses and policy makers can profit from this knowledge. Businesses can, "[…] when taking into account how their customers value their personal data, […] seek to predict which privacy-enhancing initiatives may become sources of competitive advantage and which intrusive initiatives may trigger adverse reactions" (Acquisti et al., 2013, p. 3). Policy makers can improve their analyses when implementing new strategies, if they take into account the value of privacy for their citizens. Acquisti et al. (2013) give as an example the trade-off between privacy and increased administrative cost and bureaucracy.

An interesting insight from the research of Carrascal et al. (2013) is that users valued their personal information higher when it had a direct link with their offline identity, such as age, gender, address and financial status. They suggested that a reason for this might be a lack of awareness towards the possibilities towards the tracking of online behaviour, such as which websites they visit, which friends they often email, etc. They propose that it is not immediately clear what the economic value of this might be. In our previous research we did indeed notice that users had little awareness of the possibilities of inferred data. A second interesting observation is that users do not make a distinction between the quantity of personal information that is being sold, but they do distinguish on the type. Carrascal et al. (2013) found little or no difference between the minimum amount of money they would accept for a one-time release or a 10-time release on health issues. They did however find a difference between values for types of information (financial and social information was valued higher than search or shopping information for example).

In the upcoming qualitative research, we will also take a closer look on how users value their data and their perception towards the exploitation practices of different types of information. Taking into account other types of value besides monetary, such as service-satisfaction and tailored advertising as well.

4.1.1. User Recruitment

20 Persons will be recruited for participating in ethnographic interviews, using the Q card interview technique, which will be explained in depth in the next paragraph of this deliverable. Each participant will receive a voucher of €30 for a local media store, FNAC. Each interview will take approximately 1 hour and 30 minutes and participants will be recruited in the whole region of Flanders by the living lab organization affiliated with iMinds, iLab.o.

The 20 recruited interviewees will need to have the following characteristics:

- Active Facebook users
- Computer & Internet access at home
- We will divide them on two features: low- and high-level income and age to see how this will influence their valuations of Personal Information.

Participants who are interested in staying involved in the project will be asked to also take part in the quantitative research track and open up there Facebook-data using the Databait research tool. In this way, we can link and enrich our data on their online behaviour with their claimed attitudes.

4.1.2. Q Card sorting method

The Q card sorting method has been an established way of getting insight into the subjectivity of people and their viewpoints, opinions, beliefs and attitudes towards a diversity of topics. In our research we will use the method to get insight in how people value different pieces of personal data and the reasons for their mental model. In its essence, each participant gets a stack of cards with different pieces of information on them, surrounding one topic. They are then asked to rank the cards on a certain scale, which gives the researcher insight into their subjective viewpoint. Like De Wolf & Pierson (2014), we focus on how users *could* manage and organise the different statements. Where in their research it revolved around the mapping of Facebook friends, we will use it to arrange different types of personal information.

A good overview of how this exactly works is presented by Van Exel & de Graaf (2005). We will shortly mention the five central steps they distinguished in their research and link it to our upcoming research.

1. **Definition of the concourse**

In this step all possible statements surrounding a topic are being gathered. This could be received from a number of sources, such as observation, literature, interviews, etc. In our case we will take a look at all different types of information users can disclose on their Facebook profile. This can be done by looking at the form and structure of Facebook itself, past experience of posting and the list made by our partners in deliverable 7.1.

2. **Development of the Q-set**

In the second phase the data is cleaned. After investigating all the gathered pieces a selection is made and some pieces can be put together. For example, relevant to our research, on Facebook, you can explicitly tag your location or this can be added automatically to your posts and messages. We will just make use of 'location' as a piece of information to not over complicate the matters. After a final selection is made, they are edited, randomly assigned a number and printed on separate cards (Van Exel & de Graaf, 2005)

3. **Selection of the P set**

Here the respondents are being selected. For this method it is not obligatory to get as much respondents as possible. This is driven by the idea that there only exist a number of viewpoints in society. We want to get an insight in which ones are prevalent, not on the frequency. The aim is not to indicate statistical significance or quantitative generalisability (De Wolf & Pierson, 2014). As noted by Van Exel & de Graaf (2005) "[…] All that is required are enough subjects to establish the existence of a factor for purposes of comparing one factor with another." In our research we will start with 20 persons and see if this covers all viewpoints and repetition is being noticed If not, another round of Q card sessions will be conducted.

4. **Q sorting**

In the Q sorting methods the respondents get provided with a score sheet (ranging for example from 'full disagreement' to 'full agreement' and a presentation of a quasi-normal distribution. How stronger the opinions toward the matter could be, how flatter the distribution should be drawn (leaving less room for ambiguity). In our research users for example get presented with cards on which there is always one piece of information that could be disclosed on Facebook. They can then be asked to arrange them from 'most unlikely' to 'most likely' to disclose on their Facebook profile. Because the user probably has outspoken opinions about the matter, the distribution can be made flatter, leaving more room for outspoken points of view on the outsides of the spectrum.

After the sorting is over, this can be followed by an interview, to get more insight in the whole process behind the mapping and why certain choices were made.

4.2. Research Plan – Industry Perspective

Our final research track in T4.2 contains an analysis of how personal data is valuable for the industry. More specifically, we will take a look at how different actors involved in the Flemish advertising sector currently make use of the possibilities of Personal Identifiable Information to profile users and provide tailored advertising. By enhancing our own understanding of how personal information is being used by third parties, we can provide our users with grounded information about the economic processes underlying their social relationships. In the rest of this chapter, we will present the value network that Heyman & Pierson created in 2012 containing companies that use or gather personal information in online and social media marketing campaigns. Our future research will build on this by first updating and expanding this value network in terms of number of participants and broadening the scope. To be able to do this, we will conduct expert interviews. The work done in creating the value network is also of upmost importance in relation to T3.7 of the project, where is investigated how the existing ecosystem might change under the pressure of the upcoming EU regulation and the Databait tool, when users gain more control over their personal information. For this reason the work will be aligned in the coming months.

As mentioned, the upcoming paragraph is based on the excellent work done by Heyman and Pierson (2012).

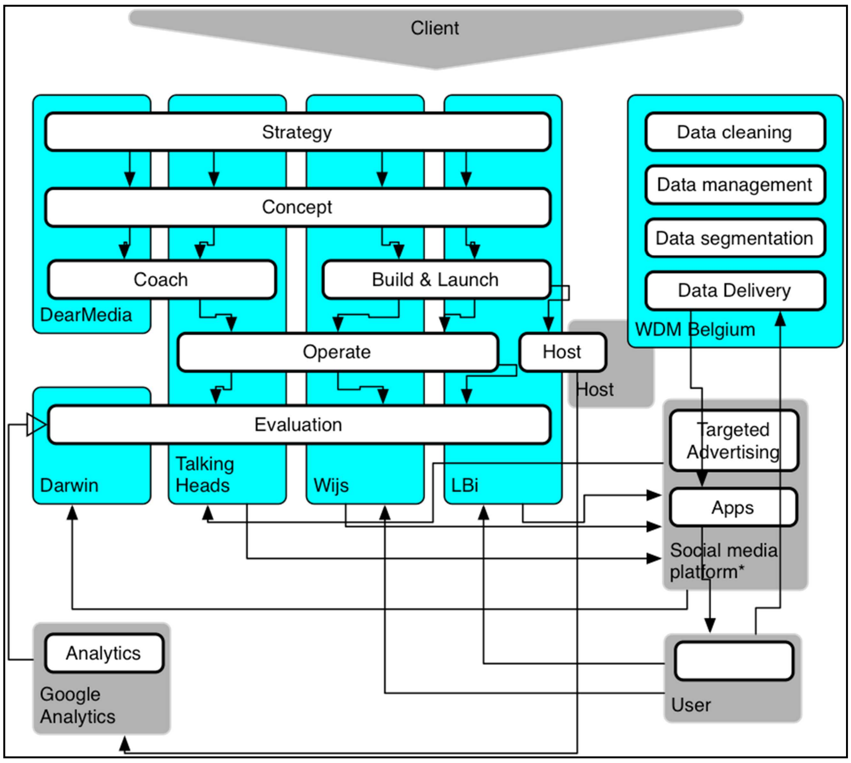


Figure 4: Value Network (Heymans, R. & Pierson (2012))

4.2.1. Presenting the Value networks

In their analyses on social media marketing campaigns in Flanders, they distinguished five actors in the value network:

- A social media consultant: responsible for designing a strategy and concept of a social media campaign
- A conversation starter: also responsible for designing a strategy and concept of a social media campaign, but also executes this process
- A digital agency: responsible for the development of an online presence through websites, and they use social media to drive users to these websites
- A social media monitoring company: responsible for measuring the performance on social media of a certain campaign
- A data supplier: uses social media to extract personal identifiable information for customer relationship management or other direct marketing purposes.

The value network itself holds the several steps and corresponding actors that a client could turn to, when wanting to develop a advertising strategy based on, or making use of social media.

The first stage they encounter is the **strategy-phase**. Here the client is usually being audited and key performance indicators (KPIs) are defined. This will guide the rolling out of the new marketing plan. Next up is the **conceptualisation stage**, here a description of what the website or social media campaign should look like is presented, keeping in mind who is being targeted and what the KPIs are. Some of the companies that were investigated by Heyman & Pierson also offer **coaching**, so their clients can learn how to execute their own social media policy. The companies involved are not so much interested in short term benefits, like increasing the number of fans on the Facebook page of the client. Instead, they want to create a long-term relationship between user and company that can result in real leads with consumer interest. Not all companies were involved in the **operation phase**, where the online marketing campaigns are rolled out to attract visitors to their webpage and social media page.

All companies use **targeted advertising** to draw users to their campaigns, where the design of the ad is data driven. The data that were most often mentioned as being valuable for this process were **gender, age, location and education**. Interesting to notice is how some of the companies prefer to lead the user as soon as possible away from Facebook. This is motivated by a number of reasons:

1. They want to avoid the Facebook censorship (e.g. the ending of a marketing campaign because Facebook’s rules are infringed)
2. When using applications, they do not want to frighten the user with the list of information they are asking. They rather believe in **progressive disclosure**: here a relationship is built and afterwards the necessary information they need is being collected on the relevant places. E.g. asking for location data when the user makes an order.

A final step is the **evaluation process** on four different and subsequent levels: raising awareness, engagement, action and finally advocacy.

To finalize our overview of the value network, we will conclude with a company that represents the data supplier mentioned above: a company that uses social media to extract personal identifiable information for customer relationship management or other direct marketing purposes. They use Facebook to gather user data, such as email addresses and basic information. They also wanted to test whether they could enrich their databases with social media data, which proved to be difficult and not very useful. Because it can be "[…] very privacy intrusive to contact a user through mail or telephone if he or she complained about services on Facebook or Twitter" (Heyman & Pierson, 2012, p. 24).

Motivation for Data Gathering in this Value Network

Two reasons are described by Heyman & Pierson (2012) on why data is being gathered in this value network:

1. To evaluate the campaigns (define ROI of campaign)
2. To identify new and existing customers

For this second motivation, the companies claimed it is interesting to get an indication of the users' budget to be able to target them with offers in their price range. This could be done by matching the newfound data with the data in their databases or to simply ask which products they like.

As seen in this overview, you notice that data gathering for tailored advertising was still limited in 2012. The main motivation for these limitations was driven by PR and economic reasons, as the companies did not want to lose potential users due to legal problems.

4.2.2. Upcoming Research

In a first step we will take a look at how the situation is now in the Flanders social marketing field and see if the presented work done by Heyman and Pierson (2012) needs to be updated. We will do this by conducting another round of expert interviews. It will be interesting to see in which way the industry has matured in the past two years with respect to data gathering and its use for social media campaigns. Another topic which is interesting to investigate is to see whether they are still primarily focused on the volunteered data on users or whether they in the meantime developed the means to infer other information that might be useful for their services, and what this information might consist of.

5.Conclusion and Next Steps

Presented in this deliverable are all steps we will take in the coming months to come up with a user defined categorisation of personal data. The literature study presented some numbers on what users generally disclose on online social network sites and how this has evolved between 2005 and 2011. Our quantitative research track, using the Databait Research Tool, provides the perfect means to build further on the presented research. It provides us with the means to perform the following tasks:

- Updating the numbers of information disclosure on online platforms, necessary in a fast-changing online media sphere.
- See how the information disclosure practices of our users evolve between 2014 and 2016.
- Validate and refine our inference algorithms to gain insights on the eight defined privacy dimensions.
- Get insights not only on volunteered data, but also observed and inferred data

Moreover, it gives us the chance to validate some insights like the effect of age on information revelation and if minorities are motivated less to disclose information on the topics that they might be discriminated against, e.g. sexual preference.

To enrich our understanding even more, a qualitative research track is proposed, with a dual perspective. With regards to the users, we want to see how they valuate the different types of personal information that they (un)knowingly share on social network sites. The proposed method was Q card sorting, as this is a great way for capturing subjectivity.

For the industry perspective we will build on the value network, which was set out by Heyman & Pierson (2012). We will do this by conducting interviews with experts in the online advertising field. Our aim is to update and broaden the suggested ecosystem in correspondence with T3.7 of the project and to receive a greater knowledge in how personal data are currently being used to tailor advertising.

6. Annex

6.1. Quantitative Research – Accompanying Questionnaire

1) **Demographics (Racial origin, Ethnicity, Literacy Level, Occupation, Income Level, Family Status)**

a) What is your **gender**?	• Male • Female • Other
b) What is your **year of birth**?	• List
c) What is your **nationality**?	• List of countries
d) What is your **country of origin**?	• List of countries
e) What is your **native language**?	• List of languages
f) What is the **highest degree** or **level of school** you have completed?	• None • Nursery school • High School • Bachelor’s degree • Master’s degree • Advanced Graduate work or Ph.D. • Not sure
g) What is your **employment status**?	• Unable to work • Retired • Student • Homemaker • Out of work, not currently looking for work • Out of work and looking for work • Self-employed • Employed for wages
h) Can you indicate on the following scale where your total household monthly income is situated?	• € 0,00 – 999,99 • € 1000,00 – 1949,99 • € 1950,00 – 2949,99 • € 2950,00 – 3949,99 • € 4000,00 or more
i) What is your **relationship status**?	• Single • In a relationship • Married • Engaged • It’s complicated • In an open relationship • Widowed • Divorced • Separated • In a domestic partnership • In a civil union

	• Something else …
j) What is your **religion**?	• Buddhist • Christian (catholic) • Christian (protestant) • Hindu • Jewish • Muslim • Sikh • Agnostic • Atheist • Other …
k) Please indicate which devices you own (select all relevant)	• Smartphone (an advanced cell phone for surfing, checking emails, using applications, etc. e.g.: iPhone, Samsung Galaxy, …) • A tablet computer (e.g.: iPad, Samsung Galaxy Tab, Asus Transformer, Microsoft Surface, …) • A desktop • A portable computer (e.g. laptop, netbook, …) • A mobile phone (only for making calls or texting) • None of the above

2) Psychological Traits

On a 7 point Likert-scale (strongly disagree – strongly agree)

I see myself as …
Extraverted, enthusiastic
Critical, quarrelsome
Dependable, self-disciplined
Anxious, easily upset
Open to new experiences, complex
Reserved, quiet
Sympathetic, warm
Disorganized, careless
Calm, emotionally stable
Conventional, uncreative
TIPI scale scoring (“R’’ denotes reverse-scored items): Extraversion: 1, 6R; Agreeableness: 2R, 7; Conscientiousness; 3, 8R; Emotional Stability: 4R, 9; Openness to Experiences: 5, 10R.

3) Sexual Profile

a) What is your **sexual preference**?	• Heterosexual • Lesbian • Gay • Bisexual • Other

4) Political Attitudes

a) Which political party has your preference? (Belgium)	• CD&V • Groen! • N-VA • Open VLD • pvda • Sp.a • Vlaams Belang • Other • None
b) Which political party has your preference? (Sweden)	• Sveriges Socialdemokratiska arbetarparti • Moderata samlingspartiet • Sverigedemokraterna • Miljöpartiet de Gröna • Centerpartiet • Vänsterpartiet • Folkpartiet Liberalerna • Kristdemokraterna • Feministiskt initiativ

5) Health Factors and Condition

a) In general, would you say that your health is:	• Excellent • Very good • Good • Fair • Poor
b) Please read all following statements carefully and tick the box next to the one that best describes you:	• I have never smoked a cigarette • I used to smoke sometimes, but I don’t now • I smoke cigarettes, but not as many as one per day • I usually smoke between 1 and 10 cigarettes per day • I smoke more than 10 cigarettes a day
c) Would you describe yourself as	• A non-drinker • A very occasional drinker (special occasions only) • An occasional drinker

	• A regular drinker
d) Please indicate next to each substance if you have used it in the last 12 months	• Coffee • Cigarettes • Alcohol • Energy drinks • Cannabis • Other drugs, …
e) Do you suffer from a chronic disease?	• Yes • If so, which one … • No

6) Location

a) In what city is your home located?	
b) In what city is your work located?	
c) What are your 5 favorite cities?	
d) Where did you spend your last holidays?	

7) Consumer Profile

Please indicate which are your hobbies/interests (you can thick as much boxes as you like)	• Reading • Watching series • Watching movies • Family time • Going to movies • Fishing • Computer • Gardening • Walking • Exercising • Listening to music • Baseball • Basketball • Ice Hockey • Soccer • Volleyball • Lacrosse • Shopping • Travelling • Sewing • Golfing • Playing music • Crafts • Watching sports • Bicycling • Playing cards

	• Hiking • Cooking • Eating out • Dating • Swimming • Camping • Skiing • Cars, motorcycles, boats • Animals • Bowling • Painting • Running • Dancing • Horseback riding • Tennis • Theater • Billiards • Beach • Other…
What are your favorite brands in the following domains?	• Cars • Clothes • Food & Beverage • Media • Music • Sports, Leisure & travel • Telecom and It • Toys • …

7.Bibliography

Acquisti, A., John, L., & Loewenstein, G. (2013). What is privacy worth? *The Journal of Legal Studies*, *42*(2), 249–274.

Bevan, J. L., Gomez, R., & Sparks, L. (2014). Disclosures about important life events on Facebook: Relationships with stress and quality of life. *Computers in Human Behavior*, *39*, 246–253. doi:10.1016/j.chb.2014.07.021

Carrascal, J. P., Riederer, C., Erramilli, V., Cherubini, M., & de Oliveira, R. (2013). Your browsing behavior for a Big Mac: Economics of Personal Information Online. *Proceedings of the 22nd International Conference on World Wide Web*, 189–200.

Christofides, E., Muise, A., & Desmarais, S. (2009). Information Disclosure and Control on Facebook: Are They Two Sides of the Same Coin or Two Different Processes? *CyberPsychology & Behavior*, *12*(3), 341–345. doi:10.1089/cpb.2008.0226

Das, S., Kramer, A., & others. (2013). Self-Censorship on Facebook. In *ICWSM*. Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/viewPDFInterstitial/6093/6350

De Wolf, R., & Pierson, J. (2014). Who's my audience again? Understanding audience management strategies for designing privacy management technologies. *Telematics and Informatics*, *31*(4), 607–616. doi:10.1016/j.tele.2013.11.004

Dwyer, C., Hiltz, S., & Passerini, K. (2007). Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace. *AMCIS 2007 Proceedings*, 339.

Feng, Y., & Xie, W. (2014). Teens' concern for privacy when using social networking sites: An analysis of socialization agents and relationships with privacy-protecting behaviors. *Computers in Human Behavior*, *33*, 153–162. doi:10.1016/j.chb.2014.01.009

Heyman, R., & Pierson, J. (2012) D3.1.2.: Analysis of business practices in profiling. EMSOC Research Report - User Empowerment ion a Social Media Culture

Hoadley, C. M., Xu, H., Lee, J. J., & Rosson, M. B. (2010). Privacy as information access and illusory control: The case of the Facebook News Feed privacy outcry. *Electronic Commerce Research and Applications*, *9*(1), 50–60. doi:10.1016/j.elerap.2009.05.001

Knijnenburg, B. P. (2014). Information Disclosure Profiles for Segmentation and Recommendation. In *Symposium on Usable Privacy and Security (SOUPS)*. Retrieved from http://cups.cs.cmu.edu/soups/2014/workshops/privacy/s3p1.pdf

Knijnenburg, B. P., Kobsa, A., & Jin, H. (2013). Dimensionality of information disclosure behavior. *International Journal of Human-Computer Studies*, *71*(12), 1144–1162.

Madden, M., Lenhart, A., Cortesi, S., Gasser, U., Duggan, M., Smith, A., & Beaton, M. (2013). Teens, social media, and privacy. *Pew Internet & American Life Project*. Retrieved from http://www.lateledipenelope.it/public/52dff2e35b812.pdf

Nosko, A., Wood, E., & Molema, S. (2010). All about me: Disclosure in online social networking profiles: The case of FACEBOOK. *Computers in Human Behavior*, *26*(3), 406–418. doi:10.1016/j.chb.2009.11.012

Pempek, T. A., Yermolayeva, Y. A., & Calvert, S. L. (2009). College students' social networking experiences on Facebook. *Journal of Applied Developmental Psychology*, *30*(3), 227–238. doi:10.1016/j.appdev.2008.12.010

Stutzman, F., Gross, R., & Acquisti, A. (2013). Silent listeners: The evolution of privacy and disclosure on facebook. *Journal of Privacy and Confidentiality*, *4*(2), 2.

Taraszow, T., Aristodemou, E., Shitta, G., Laouris, Y., & Arsoy, A. (2010). Disclosure of personal and contact information by young people in social networking sites: An analysis using Facebook profiles as an example. *International Journal of Media & Cultural Politics*, *6*(1), 81–101. doi:10.1386/macp.6.1.81/1

Van Exel, J., & de Graaf, G. (2005). Q methodology: A sneak preview. *Online Document. Http://www. Qmethodology. net/PDF/Q-Methodology*. Retrieved from http://qmethod.org/articles/vanExel.pdf

World Economic Forum. (2012). *Rethinking Personal Data: Strengthening Trust* (p. 36). Switzerland.

Young, A. L., & Quan-Haase, A. (2009). Information revelation and internet privacy concerns on social network sites: a case study of facebook. In *Proceedings of* *the fourth international conference on Communities and technologies* (pp. 265–274). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1556499