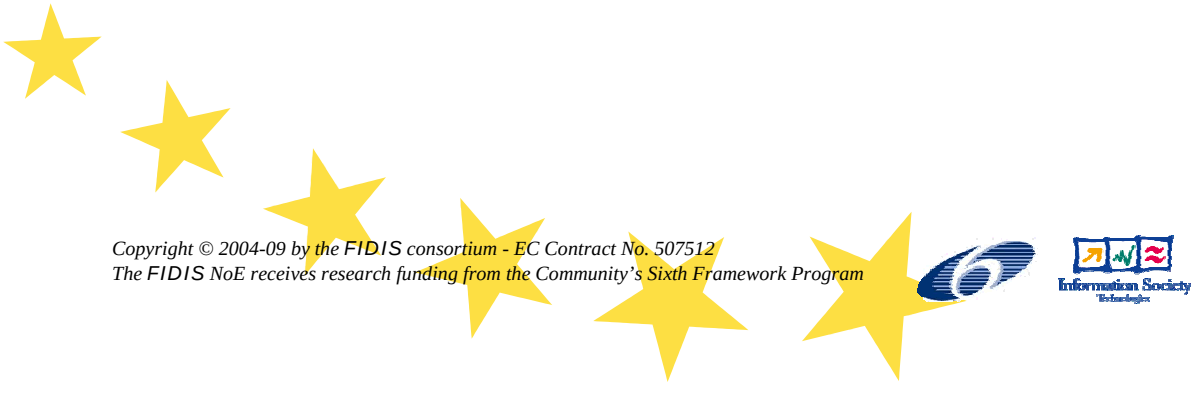
FIDIS

Future of Identity in the Information Society

Title:	"D14.5 Experimental Study on Profiling in Business Processes"
Author:	WP14
Editors:	Rani Husseiki (SIRRIX)
Reviewers:	Uli Pinsdorf (Microsoft), Zeno Gerardts (DT)
Identifier:	D14.5
Type:	[Deliverable]
Version:	1.4
Date:	Friday, 26 June 2009
Status:	[Final]
Class:	[Public]
File:	fidis_wp14_d14.5_v1.4.doc

Summary

The aim of this study is in tracing the behaviour of mainly commercial entities with respect to their handling of personal data. Many profiling activities are done without a clear legal base: personal data is passed through without the explicit consent of the data subjects. The experiment is an empirical analysis, giving an understanding how companies and authorities are dealing with personal data. The study is conducted as a filed study were personal data is marked (e.g., by slightly modifying or misspelling personal data) and given away to commercial companies (e.g., buying portals, club cards etc.). The study is a mid-term study. Based on the received postal and electronic advertisements, it can be traced which entities has leaked personal data.



The FIDIS NoE receives research funding from the Community's Sixth Framework Program

Copyright Notice:

PLEASE NOTE: This document may change without notice – Updated versions of this document can be found at the FIDIS NoE website at www.fidis.net.

Members of the FIDIS consortium

1. Goethe University Frankfurt	Germany
2. Joint Research Centre (JRC)	Spain
3. Vrije Universiteit Brussel	Belgium
4. Unabhängiges Landeszentrum für Datenschutz (ICPP)	Germany
5. Institut Europeen D'Administration Des Affaires (INSEAD)	France
6. University of Reading	United Kingdom
7. Katholieke Universiteit Leuven	Belgium
8. Tilburg University[1]	Netherlands
9. Karlstads University	Sweden
10. Technische Universität Berlin	Germany
11. Technische Universität Dresden	Germany
12. Albert-Ludwig-University Freiburg	Germany
13. Masarykova universita v Brne (MU)	Czech Republic
14. VaF Bratislava	Slovakia
15. London School of Economics and Political Science (LSE)	United Kingdom
16. Budapest University of Technology and Economics (ISTRI)	Hungary
17. IBM Research GmbH	Switzerland
18. Centre Technique de la Gendarmerie Nationale (CTGN)	France
19. Netherlands Forensic Institute (NFI)[2]	Netherlands
20. Virtual Identity and Privacy Research Center (VIP)[3]	Switzerland
21. Europäisches Microsoft Innovations Center GmbH (EMIC)	Germany
22. Institute of Communication and Computer Systems (ICCS)	Greece
23. AXSionics AG	Switzerland
24. SIRRIX AG Security Technologies	Germany

[1] Legal name: Stichting Katholieke Universiteit Brabant
[2] Legal name: Ministerie Van Justitie
[3] Legal name: Berner Fachhochschule

Versions

Version	*Date*	*Description (Editor)*
1.0	February 26	• Table of contents (Rani Husseiki)
1.1	March 16	• Results collection (Rani Husseiki)
1.2	April 29	• Review version (Rani Husseiki)
1.3	June 15	• Second Review version (Rani Husseiki)
1.4	June 25	• Final version (Rani Husseiki)

Foreword

FIDIS partners from various disciplines have contributed as authors to this document. The following list names the main contributors for the chapters of this document:

Chapter	*Contributor(s)*
1 Executive Summary	Rani Husseiki
2 Introduction	Rani Husseiki
3 Profiling in Business Processes	Rani Husseiki
4 Existing Studies on Profiling	Rani Husseiki
5 Experimental Study	Rani Husseiki
6 Analysis	Rani Husseiki
7 Conclusion	Rani Husseiki
8 Biography	Rani Husseiki

Table of Contents

1 Executive Summary

This document describes an empirical study that complements the FIDIS studies that have been focused on profiling and privacy issues from a theoretical perspective. It shows another perspective of profiling and privacy breaches in business processes based on collected evidence by the normal customer.

The experiment pivoted around the idea of name-watermarking and targeted institutions that typically allow registration of customers by providing personal data. The goal was to register at a variety of business institutions and wait until post and email letters are received from these institutions in order to derive conclusions of privacy breaches and profiling activities that are typically carried out by collecting profiles of customers and leaking them to partner businesses. The name-watermarking method aimed at tracing the process of leaking profiles to other companies in order to obtain profiling traces that can be used as an indicator, at least for the customer himself, of misuse of private information by a specific company.

The experiment spanned over a period of three months and a half. During this time, we collected post and email from the businesses that we already registered at using *derived identities*. We categorized the results, mainly according to three factors, which are:

- The pre-existence of a privacy policy that has been agreed on between the target institution and the customer upon registration to the service.
- The proof of compliance or no compliance to such a policy, if it exists.
- The identification of the result as a profiling trace based on the substantial proof that the profile has been leaked to another institution.

The results of the experiment were analysed, which led to the following observations and conclusions:

1) Profiling activities as well as privacy violations are not as frequent as we expected. Therefore, it is relatively hard for a normal customer to perceive privacy breaches or profiling behaviour by the companies he is registered at.
2) Customer loyalty programs (or bonus programs) are generally not inclined to violate privacy of their customer's data, at least not more than other type of companies (e.g. in the scope of customer-relationship management activities). Customers are therefore generally unable to perceive such indication, which contradicts our original hypothesis. However, they might carry out profiling activities customers in a discrete manner which is not perceivable to customers.
3) A proof of misuse of personal data by companies often can not be regarded as a legal proof of privacy breaching either because the behaviour is still within the boundaries of a privacy policy agreed on with the customer, or because an agreement on such a privacy policy has been initially avoided by the company.

Nevertheless, an assessment of the scientific value of the conducted field experiment was given, which led to the conclusion that many methodological aspects lack accuracy or adequacy. However, the interpretation approach that led to our observations was fairly acceptable. Next steps were proposed as recommendations for an experiment of "more" scientific nature, and techniques to combat profiling activities and privacy violations.

2 Introduction

Personalized services are often adopted by private and public organizations and institutions. Statistical studies show that the burden of filling out forms with personal information by citizens can be impeding the normal administrative and business processes. Especially, this requirement seems to be annoying when customers, citizens or users find themselves filling out forms with the same personal data that they have provided over and over again.

Particularly in business processes, services and products are being more and more personalized, based on customer data that is collected and saved into user profiles. In D14.3, the business processes requiring personal data have been explained, including the data collection or information flow model, a use case on personalized services, and the trust model concerning processing of personal data.

It is self-evident that this data collection model poses privacy concerns to costumers regardless of the advantage provided by the personalized services built on it. The requirements for verifiable data processing have been derived in D14.3, and technical measures for providing policy-compliant data processing by service providers have been proposed. Those requirements and measures are based on legitimate concerns from the customers or contributors to business processes, especially when they notice that their personal data has been somehow leaked to some advertising company.

In this deliverable, we aim at setting up a study based on tracing the behaviour of commercial entities with respect to handling personal data.

In fact, profiling activities have thoroughly been addressed in many FIDIS deliverables, particularly in Del 7.2 on “Descriptive analysis and inventory of profiling practices”. While those deliverables have discussed the issues from a conceptual perspective, it is worth attempting to provide a complementary study of empirical nature that allows supporting the theoretical results by substantial proofs, in addition to a reflection on them. This enables us to see what the single customer would normally perceive out of these profiling activities and privacy breaches with respect to the personal data he usually provides to business companies. It also helps us to understand what kind of the proofs can be collected that be used as a legal basis for uncovering misuse of personal data if privacy measures are to be adopted against that.

Many profiling activities are done without a clear legal base: personal data is passed through without the explicit consent of the individuals. Therefore, an empirical analysis can be very effective towards giving an understanding on how companies and authorities are dealing with personal data.

This experimental study is conducted as a field study were personal data is marked (e.g., by slightly modifying names, data etc.) and given away to commercial companies (e.g., buying portals, club cards etc.). Based on the received postal and electronic advertisements, the analysis uncovers which institutions have leaked the personal data, which ones have violated privacy agreements with their own customers, and which ones have taken advantage of the customer’s inexperience in legal matters in order to misuse his private information against his will or consent.

3 Profiling in Business Processes

Profiling activities can take several forms and might have different purposes. In this deliverable, we focus on profiling in business processes, particularly profiling activities which are carried out by business institutions on customers of certain services they provide. While the study is of an experimental nature, it is important to shed a light on the how *profiles* and *profiling* are generally defined, in which business scenarios they are typically encountered, and what the legal implications of profiling by business partners are.

3.1 Definitions of Profiles and Profiling

Literature reviews, as well as dedicated studies about profiling ([Hild08], [Del 7.2]), clearly indicate that the term represents a complex concept. It follows that multiple definitions of the term can be encountered across the literature, but there are common aspects for the term that are widely agreed on.

Before defining what profiling is, it is important to give a short overview of typical profiles of users that are usually constructed within business processes.

In deliverable Del 7.2 (Descriptive analysis and inventory of profiling practices, section 2.3), a profile is defined as "a set of correlated data that identify and represent a data subject". "When the data subject is a single person we speak of a personalised profile, when the data subject is a group/a category or a cluster we speak of a group profile."

In business processes scenarios (e.g. deliverable Del 14.3 section 3.3 on personalized data), this set of personal data that identify the data subject – or service user in this case – is constructed mainly out of the personal information provided by the user upon registration to the service (whether digitally or via paper forms), in addition to some user-specific usage patterns of the service which becomes part of the corresponding user's profile.

Those comprehensive profiles are therefore formed through information collected by the company either with or without the consent or knowledge of the individual using the service. In many cases, the user is not given notice that certain information is being collected about him, or not given the choice to decide which information he prefers to opt-out of the profile.

In many situations, these profiles are used by the marketing department within the company for advertising purposes [Epic[4]]. Sometimes, companies go as far as selling collections of profiles to other companies that wish to perform market analysis or carry on their own advertising activities. The user becomes a target of direct-marketing campaigns by means of spamming, or psychological manipulation through targeted advertisement.

Profiling has been addressed in many deliverables. One definition of profiling that can be found in Del 7.2 is:

a. the process of constructing profiles (correlated data), that identify and represent either a person or a group/category/cluster,

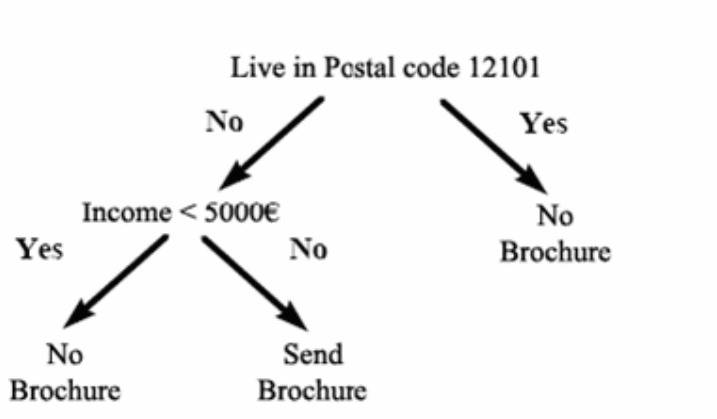
[4] Epic http://epic.org/privacy/profiling/

b. and/or the application of profiles (correlated data) to identify and represent a person as a specific person or as member of a specific group/category/cluster; (D7.2, 2.8)

c. and/or aiming at the assessment of risks and/or opportunities for the data user (inferred from risks and opportunities concerning the data subject).

3.2 Profiling Scenarios

An example profiling scenario for the advertisement purposes mentioned in D7.2 (3.2.3.4) is depicted in the figure below. Customers registered for a certain service become subject to advertisement activities through brochures based on attributes collected about them as part of their profiles. In this particular scenario, the "address" attribute and the "income" attribute in each customer profile are used to categorize customers for a brochure mailing.



The attributes based on which the profiles have either been provided by the customers themselves (e.g. postal code), other attributes (e.g. income) might be obtained from another source without the knowledge of the customer. Personalized profiles collected from other companies are such an attribute source.
Targeted advertisements based on customer profiles are known to be more effective.

The particular profiling scenario we focus on in this deliverable is the profiling activities carried out by in the context of customer loyalty programs[5] which are meant for discount purposes. In this type of business processes, personal information about a customer is initially obtained upon enrolment of the customer to the program. According to Del 7.2 (section 5.1.1), in most cases, additional personal data that might not be needed for the discount operation are collected. Those include:

[5] "Loyalty programs are structured marketing efforts that reward, and therefore encourage, loyal buying behavior — behavior which is potentially of benefit to the firm. In marketing generally and in retailing more specifically, a loyalty card, rewards card, points card, advantage card, or club card is a plastic or paper card, visually similar to a credit card or debit card, that identifies the card holder as a member in a loyalty program." – Wikipedia.

- Date of birth
- Multiple contact addresses (telephone numbers, e-mail etc.)
- Which goods where purchased when and where
- Information on personal circumstances of life (e.g. family status, number of children, income etc.)

Based on this data, profiling activities are practiced for market research and advertising purposes.

3.3 Legal aspects

Many legal aspects are related to profiling activities since they typically involve handling of personal data which entail privacy requirements.

If we consider the example of online personalization, where a certain service is tailored to the needs, interests, and convenience of a certain customer, we can derive four main privacy aspects (Del 7.2, 4.3.1):

1. A service-provider approach through which requested user information is not strictly related to the delivery and access of a specific service;

2. User-data collection using invisible methods, which use spy technologies, such as cookies, web bugs, etc. to trace, track and search user profiles;

3. Use of personal data for purposes different from those indicated and without the user's previous and/or informed consent;

4. Lack of effective user access to personal data collected, e.g. at web sites.

The experiment carried out in the context of this deliverable focuses on the 1st and 3rd aspect. Particularly, customer loyalty programs are targeted in order to inspect how the corresponding institutions can use the personal data for purposes which are different from the ones indicated in their privacy policy which represents the only legal ground for the handling of the personal data.

In appendix B of Del 7.2, the legal grounds for customer loyalty programs are mentioned. While our experiment does not solely focus on this type of business, customer loyalty programs are more relevant for such an experiment since they are sometimes regarded to be more prone to using personal data for their own interests [Dunn, 2005], e.g. marketing purposes in order to account for the benefits they are providing to the customer.

Particularly, in [ICPP] profiling in customer loyalty programs was addressed, and based on German legislations by BDSG[6] (which has several sections applicable to customer loyalty programs) it defined the type of personal data that are allowed to be collected and processed for this kind of programs are:

[6] German Federal Data Protection Act

- Name
- Address
- Year of birth
- One further contact information (phone, email etc.)
- Time and place of card deployment
- Price of purchased goods / services and discount amount
- Data related to the purchased goods only if they are necessary for computation of the discount amount.

Further legal implications of specific articles of the BDSG on the misuse of personal data and the rights of the customer in customer loyalty programs can be found in Del 7.2 Appendix B.

4 Existing Studies on Profiling

Prior to going in the details of the experimental study carried out by Sirrix, we give a brief overview of example existing studies on profiling and their general results and conclusions.

4.1 Theoretical Studies

Theoretical studies on profiling have different methodologies. They might include an evaluation of available identity management and data mining systems in terms of how they handle personal data. Another method would be to notice general patterns of profiling activities that are, e.g., encountered by consumers and employees, and attempt to draw conclusions on the means, types and implications of profiling. As opposed to an empirical study, a typical theoretical study would not include a systematic approach for collecting profiling traces as perceived by the citizens or consumers.

One example of theoretical studies on profiling has been mentioned in Del 7.2, section 5.1.1. This study on 16 customer loyalty programs has been carried out by ICPP in 2003 and results were evaluated against the German Federal Data Protection Act (BDSG). According to this study, "a central weakness of all investigated programs is that because of trade secrecy, the place and time of storage of the data, the way and the purpose of processing was not described sufficiently. On the grounds of insufficient information, a declaration of consent - which has to be based on a free will - is legally not effective."

[Hild2] presents a chapter with an attempt to define profiling, where different types and means of profiling are distinguished, for example between individual profiles and group profiles. Machine profiling constituted one focus of the study. One conclusion from the work is that consumers may find themselves "in the position of being profiled, without access to the knowledge that is used to categorise and deal with them".

These studies constituted a motivation for conducting an experimental study that would show how consumers can actually perceive profiling if a systematic approach for collecting traces and analysing them is followed.

4.2 Empirical Studies

Empirical studies on violations of organizational security policies have also been carried out before.

For example, a study [CIPP1] on privacy violations in organizations carried out in 2006 by the Canadian Internet Policy and Public Interest Clinic (CIPPIC) on 64 online retailers in order to assess their compliance with privacy laws. The results showed that a "surprising number" of companies failed to comply with the rules. The following is a summary of the aspects behind ineffective compliance to privacy laws by organizations:

- Difficulty of consumers to get answers regarding data protection policies of the company.

- Parts of privacy policies are unclear.
- Policies are often incomplete (unclear agreements with TTP).
- Notices of secondary usage of private data are often not included in the policy.
- No choice given to customers to decide on unnecessary usage or disclosure of his data.
- Policy terms on sharing private data only with customer's consent are often disregarded.

Another study [CIPP2] by CIPPIC entitled “the Data Trail: How Detailed Information About You Gets Into The Hands Of Organizations With Whom You Have No Relationship” shows how customer private data is gathered and traded in the marketplace by organizations collecting information from customers. The study suggests that many organizations consider the benefits of sharing private customer data to outweigh costs. For example, private customer information can often be traded directly between data owners and users. The study showed that most consumer data is offered for rent or sale in the form of a list of names and credentials including telephone numbers, email addresses, postal addresses, etc. Sources of consumer data were found to be various retailers and service providers such as magazines, newspapers, email and other subscription services, travel agencies, product manufacturers (via registration/warranty cards), online educational and information services, and payment processing companies.

5 Experimental Study

In our experimental study, we shed the light on a critical aspect of profiling in business processes based on empirical data reflecting personal data mishandling and privacy violations for the purpose of profiling. The study addresses the particular profiling trend in organizations. It aims at collecting substantial proofs to trace the behaviour of commercial entities with respect to their handling of personal data. Basically, the experiments try to uncover the fact that personal data is passed through to other companies, or misused by the original company itself without the consent of the corresponding customers. When this mishandling of personal data is done based on the particular activity or preference of the customer, a profiling trace is declared.

While the experiment targets a wide range of institutions, we lay down the hypothesis that customer loyalty programs are more inclined to violate the privacy of their customers' personal data or to carry out profiling activities on this data as compared to other institutions. The reason is, we think that these programs (also called bonus programs) have more interest in sending out tailored advertisements to their customers, or leaking their profiles to other institutions in the sake of a financial gain that would compensate for the granted bonuses. This is, however, just a hypothesis that the experiment might either prove or refute.

The extent of privacy violation by organizations collecting *personally identifiable information* can hardly be estimated unless the evaluation of privacy breaches is done with respect to a privacy policy which is previously defined and agreed on between the organization collecting the information and its customers. Therefore, in order to assess and eventually limit the extent of privacy violations in organizations the following should be addressed:

1) Pre-evaluation of the reliability of an organization's privacy policy, for example, by studying and evaluating the procedural and technical measures which are supported by the organization in order to enforce the privacy policy within its business processes.
2) Post-evaluation of the reliability of and the compliance with the privacy policy, for example, by evaluating and assessing the privacy breaches that occur – whether intentionally or unintentionally – with respect to the terms of the privacy policy.

While the first aspect is important and reflects a proactive approach towards establishing an organizational strategy for ensuring privacy in business processes, the second aspect helps the organization figure out the weak points in the procedural and technical measures applied. On the other hand, the customers are able to assess and judge on the compliance of the organization to the privacy policy.

Since our study relies on empirical data, it falls in the second category of evaluation. The methodology of the study is a form of post-evaluation of the compliance of organizations to privacy policies by means of experimental analysis. Personal data is marked (e.g., by slightly modifying names, data etc.) and given away to commercial companies (e.g., buying portals, club cards etc.) with the purpose of tracing the leakage of personal customer information based on, e.g., advertisements.

In the following sections, we elaborate on the details and results of the experiment.

5.1 Target Institutions

The experiment does not take into consideration all the types of privacy breaches on personal data that might occur within organizations. It rather focuses on business institutions that often require that a user or customer to the business provides identity information to the business.

While it is self-evident that not all businesses are equally trustworthy to keep the identity data confidential, or does not abide by privacy policies, we aim at proving this belief by means of concrete proofs. The results of the experiment would show how certain privacy policies are being breached, whether by misuse of the personal data by the same company to which the data has been granted, or by sharing the identity information with other business partners.

The idea is to register with several businesses with distinguishable and unique identities. Then any postal or electronic mail received from those businesses or other businesses will be identified and checked in order to draw conclusions around the general ways and circumstances of privacy breaches in businesses.

5.2 Methodology

First, a set of fourty-six institutions were chosen as target institutions. Those were chosen from a wide range of fields, namely Electricity, Health insurance, Mobile phone operators, Internet providers, Catalogues, Job search engines, Social networking, Clothing stores, Book stores, Auction systems, Food stores, Cosmetics stores, Flight operators, Email websites, Newspapers, Magazines, and Bonus programs (see Table 1). The purpose was to cover a wide range of institutions, which would give a better insight on the privacy breaches.

A relatively long, non-western full name (BHARATANATYAM PRIYADARSHANI) was chosen along with identity attributes, and was used to derive different identities (see 5.4). The original name was printed on the post box of Sirrix (the company conducting the study), and a set of personal data (address of the company, email address, birth date, etc…) attributed to the name was created to form a full identity.

Fifty derived identities were created by introducing variations to the name, but leaving the identity attributes common to all names. Then, registration has been performed in each of the fifty chosen institutions, each with a distinct derived identity. So each derived identity is handed over to only a single business. This means that each company had a slightly different name used for registration, but the post address and email address were the same for all profiles. The mapping between Derived ID and Institution has been stored for correlation between received post or email and the sending institution on a later stage.

The chosen companies are all located in Germany. During this time, the fictious name was left on the post box, and post as well as emails were collected and categorized.

The experiment was conducted between September 2008 and March 2009. The period allocated for the experiment was extended from 4 months (originally) to 6 months, in order to allow for more traces to be collected since we noticed that the rate of mails and emails received was considerably low.

5.3 Chosen Institutions

In order to make the range of chosen companies as wide as possible, we defined a set of categories of companies. However, we chose companies to be in Germany in order to be able to compare the results to a single legislation. It also guaranteed a higher number of post received, and therefore a better possibility to obtain traces of privacy breaching and profiling activities. The general categories were:

- "Basic life services" which included telephone, electricity, insurance, life insurance, banks, mobile companies and internet providers.
- "Online shops": clothes, books, auction systems, food, logistics, furniture, cosmetics, flight operators, email, lottery, pay TV, newspapers, magazines.
- "Miscellaneous": catalogue sending, job search engines, social networking websites.
- "Bonus programs": gas stations points, flight operator miles, shopping cards.

Company Category	Targeted	Company Category	Targeted
Electricity	2	Auction systems	2
Health insurance	3	Food stores	2
Mobile companies	2	Cosmetics stores	4
Internet providers	2	Flight operators	2
Catalogue	2	Email websites	3
Job search engines	4	Newspapers	4
Social networking	2	Magazines	2
Clothing stores	3	Bonus programs	5
Book stores	2		

Table 1: number of target institutions per field.

The targeted businesses are listed below according to the categories they belong to:

Electricity	**Job Search Engines**
o RWE	o Monster
o EON	o Jobpilot
Health Insurance	o Jobs.de
o AOK	o Stepstone
o DAK	**Social Networking**
o IKK	o StudiVZ

Mobile	o WerKenntWen
o O2	**Clothing stores**
o T-COM	o S. Oliver
Internet	o H&M
o 1&1	o Orsay
o Freenet	**Book stores**
Various Outfits	o Amazon
o Quelle	o Mayer
o Reichelt	**Email**
Food	o Web.de
o TCHIBO	o Gmx
o PLUS	o Yahoo
Cosmetics stores	**Newspapers**
o DM	o Frankfurter Allgemeine
o Schlecker	o Süddeutsche Zeitung
o Rossmann	o WAZ
o Douglas	o BILD
Flight Operators	**Bonus programs**
o Lufthansa	o Payback
o German Wings	o Lufthansa Miles&More
Magazines	o IKEA family card
o Spiegel	o Shell club smart
o Focus	o Happy Digits
Auction Systems	**Electronics**
o Ebay	o Saturn
o	o Conrad Electronics

Table2: targeted institutions per category

5.4 Traceable Profiles

The idea employed in this study is to use tracing methods to find how data is misused, and where data flows, when handed over to a business that is suspected to misuse it or share it. While finding proof of unauthorized data misuse or sharing is certainly worthwhile and

rewarding, assessing the types of privacy breach and even finding the concrete proofs of profiling or sharing of identity data gives insight into patterns of mishandling of this identity data.

Generally speaking, suspected data sharing can be traced either internally to a business or external to it. We explain in the following what these general terms shall mean with respect to this study.

Usually, the internal tracing of where data flows gives the best and reliable results. However, examination of processes and procedures internal to a business are typically not available to outsiders, as they are often considered a trade or business secret.

External methods of tracing data sharing or data leakage try to find data that is supposed to be available only to business A, but appears to be known to other entities as well. To achieve results with such methods, there needs to be a way to detect the shared data. That means, that the data needs to be in a specific form that is given to the suspected data sharer in exclusive form. In a way this resembles the concept of watermarking data in order to later find copies and correlate the data found to the entities the data was given to.

While in recent years powerful and robust watermarking schemes have been developed for image, video, and audio data, we are much more constrained here when it comes to watermarking identities.

Further, the robustness of watermarking identities might be significantly lower than the robustness of established schemes for image, video, or audio data. In fact, once several instances of the same identity in watermarked form can be correlated, the watermarks might be removed, or made unintelligible.

Nevertheless, we discuss next a simple name-based watermarking scheme for identities that we deem sufficient for the purpose of conducting this experimental study.

A straight-forward scheme for watermarking identities is based on the idea to produce several identities from a real one. This is done by introducing slight variations that could also appear as data errors, such as placing different middle initials, variations in the spelling of non-western names, permuted characters or other intentional minor changes. Denote such an identity that is based on a real identity by introducing a variation a *derived identity*.

Here, the correlation between a specific derived identity and a specific business can easily be made by storing the fact that the derived identity was handed over to the business, while making sure that each derived identity is handed over to only a single business.

However, when implementing such a watermarking in practice, care must be taken when it comes to how large a variation can be, and especially for identity data that is used electronically, such as email addresses. An example for the former is the mismatch between a name on a letter and the name on the post box for physical mail. Regarding the latter issue, it must be ensured that for each derived identity a different email address is used. However, several email addresses in the same domain might be used.

We chose a long non-western name in allow more possibilities for unrecognizable permutations. We started by the following set of identity information:

```
First name/ Last name: BHARATANATYAM PRIYADARSHANI
Date of birth: 21.08.1973
Address: Lise-Meitner-Allee 4; 44801 Bochum; Germany
```

Telephone: according to company

bhar.priy@yahoo.com

Based on a systematic approach for permutations, we could obtain around fifty *derived identities*. This was done by systematically permuting vowel letters in the first and last name. The permutations of vowels in the first and last name were done in a way so as to derive unique full names. The other attributes were used without modifications for all the target businesses. The variations have been introduced in a way that is limited enough to appear as a spelling mistake (a typo) rather than an intentional variation or spelling of a different name. Each derived identity was associated with a target company, and the associations were saved.

The following are few examples of derived identities:

BHARATANATYAM PRIYADARSHANI	Original
BHAREТANATYAM PRIYADIRSHANI	Lufthansa Miles&More
BHARATENATYAM PRIYADARSHINI	IKEA Family card
BHARATANETYAM PRIYEDARSHANI	Douglas Card
BHARATANATYEM PRIYADERSHANI	Shell smart card

Only the name was watermarked, the other attributes were left as they are for all derived identities. Particularly, the email was not watermarked and was common between all identities. Otherwise, we would have had to create 46 different email accounts and monitor them, which is a cumbersome approach that does not substantially differ in the tracing process. Our results proved that all received mails and emails contained the name of the addressee which was enough as a trace for the company who sent the post/email or leaked the targeted profile.

5.5 Results

The results that were received are provided below. For each received post or email, the category of the original company is noted (e.g. online shop, bonus program), as well as the existence of a prior privacy policy agreement with the customer, and whether this policy has been respected or not. The decision on whether a profiling trace is found is based on the fact that the post or email has been received from a company that has not been provided any identity information. The results were categorized as follows:

- **Result**: the actual item received, e.g. newsletter, promotions booklet, discount or advertisement brochure, etc…
- **Result Source**: the source of the result i.e. the original company were registration has been made, or some other company to which the profile has been obviously leaked.
- **Result Type**: the means by which the result was received i.e. Post or Email.
- **Company Type**: the general type or category of the company as per Table 2.

- **Policy Agreement with customer**: this field indicates whether a privacy policy has been agreed on with the customer during registration to the service (e.g. by asking the customer to confirm having read the terms and conditions of the policy).
- **Policy Compliance**: this field indicates whether the received result (email or post) violates the privacy policy that has been agreed on with the customer. In case no privacy policy exists, we use N/A.
- **Privacy Breach**: a privacy breach is declared whenever the result has been sent from a company which is different from the company at which the customer originally registered (this indicates that the profile of the customer has been leaked for some reason). Or when there has been a violation of a privacy policy which is previously agreed on.
- **Profiling Trace**: a profiling trace is declared when the received results seems to be addressed to this particular customer based on a previous usage of the service by the customer. Or when the result source is different from the original company, but still is from the company type. In both cases, the customer is believed to have been particularly targeted based on his interest.

Result	**Result Source**	**Result Type**	**Company Type**	**Policy Agreement w/ customer**	**Policy Compliance**	**Privacy Breach**	**Profiling Trace**
Newsletter	Original Company	Email	Bonus Program	Yes	Yes	No	No
Newsletter	Original Company	Email	Online Shop	Yes	Yes	No	No
Promotion	Original Company	Email	Online Shop	No	N/A	No	Yes[1]
Newsletter	Original Company	Email	Bonus Program	Yes	Yes	No	No
New Items Advertisement	Original Company	Email	Online Bookshop	Yes	Yes	No	Yes[2]
Newsletter	Original Company	Email	Bonus Program	Yes	Yes	No	No
Discounts Advertisement	Original Company	Post	Clothing Shop	No	N/A	No	No
Order Magazine	Original Company	Post	Various Outfits Shop	No	N/A	No	No
Promotions	Original Company	Email	Online Clothing Shop	No	N/A	No	No
Promotions	Other Company	Email	Online Electronics Shop	No	N/A	Yes	Yes[3]
Promotions	Original Company	Email	Social Networking	Yes	No	Yes	No
Promotions	Other Company	Post	Newspaper	No	N/A	Yes	No

Promotions	Original Company	Post	Cosmetics	No	N/A	No	No
Newsletter	Original Company	Email	Bonus Program	Yes	Yes	No	No
Newsletter	Original Company	Email	Bonus Program	Yes	Yes	No	No
Newsletter	Original Company	Email	Bonus Program	Yes	Yes	No	No

Table 5.5: Categorized Received Post or Email

[1] The received promotion was about the same particular category of goods that has been bought by the customer. (we actually bought one item through this online shop using the derived identity).

[2] The received advertisement was about the same particular category of books that has been searched by the customer (we added a book to the shopping basket without buying it).

[3] The received promotion was from another company than the original company, but was about electronic devices, which reflect the original company's type.

6 Analysis

In the following, we list some observations about the results obtained based on Table 5.5, and draw conclusions on privacy breaches and profiling by the different companies.

6.1 Form of results

The results obtained where in three forms, namely:

- Promotions (6 out of 16)
- Newsletters (7 out of 16)
- New Services / Discounts advertisements (3 out of 16)

The results show that all received post or email had the purpose of advertisement, whether for the company to which the personal information has been granted, or the company to which this information has been leaked.

6.2 Privacy Violations, and Profiling Traces

One a first stage, the analysis of the results led to the following observation: the type of email/post received can be categorized into four cases:

A. **Legitimate:** this is when neither a privacy breach nor a profiling trace could be found. The advertisement does not seem to be addressed to this customer based on particular aspects of his profile. The customer has agreed on a privacy policy on identity information usage which was proposed by the company at the registration time and that the advertisement does not violate (Yes, Yes, No, No). Or he did not agree on any privacy policy a-priori (No, N/A, No, No).

B. No Profiling, Privacy Violation: this is when a privacy violation has occurred without any indication that the advertisement was addressed to this customer based on particular aspects of his profile. However, a privacy policy on identity information usage was proposed by the company at the registration time, and the customer has supposedly agreed on, but the advertisement received does not comply with the policy, which means that the company has misused the identity information without actually leaking it. (Yes, No, Yes, No). Another case would be if no agreement on a privacy policy exists, but the source of the advertisement is a company which is different from the original company, which shows that the profile has been leaked (No, N/A, Yes, No).

C. Profiling, No Privacy Violation: this is when a profiling trace is detected even if no privacy violation has occurred. This means that the company complied with its privacy policy (Yes, Yes, No, Yes), or no privacy policy was agreed on a priori (No, N/A, No, Yes). However, the advertisement seems to be addressed to this particular customer based on his previous activity (e.g. the advertisement is about items or goods that relate to a previously purchased or searched item).

D. Profiling, Privacy Violation: in this case, the privacy of the customer has been violated, probably based on his profile. E.g. his profile seem to have been leaked to another company, which sends advertisements about the same goods of the original company. This indicates a profiling trace (No, N/A, Yes, Yes).

6.2.1 Category A (Legitimate Activity)

In this category, fell 5 results out of 16 received results. The following types of companies:

- 4 Bonus Program (out of 5 registered at, with two redundancies)
- 2 Clothing store (out of 3 registered at)
- 1 cosmetics (out of 4 registered at)
- 1 food store (out of 2 registered at)
- 1 various outfits shop (out of 2 registered at)

6.2.2 Category B (No Profiling, Privacy Violation)

In this category fell only 2 out of 16 received results. The type of the company is:

- 1 Social Networking (out of 2 registered at)
- 1 Newspaper (out of 4 registered at)

6.2.3 Category C (Profiling, No Privacy Violation)

In this category fell 2 out of 16 results received. The types of companies are:

- 1 Online clothing shop (out of 3 registered at)
- 1 Online Bookshop (out of 2 registered at)

6.2.4 Category D (Profiling and Privacy Violation)

In this category, fell 1 out of 16 results received. The types of companies are as follows:

- 1 electronics shop (out of 2 registered at)

6.3 Observations

While it might not be accurate to make statistically based conclusions on the obtained results due to the fact that the number of results is limited, few observations must be noted:

1. The number of results obtained over the 4 months period of time was considerably less than expected (see hypothesis in 5) at the beginning of the experiment. This showed that despite the fact that many hypothetical analyses come to the conclusion that profiling activities and privacy breaches by institutions with respect to personal data are at a high level, it might be hard for a single customer to perceive those activities if he is registered at only a few of these companies. However, it should be noted that a single breach of a privacy policy through an advertisement, or a single proof of leakage of personal data to another institution, can be enough evidence for a customer to confirm his original expectations.
2. Profiling traces were considerably low among the results received (only 3 out of 16 results), but still existed. At least from the results received, it seems that if an institution allows itself to misuse the personal information of a customer, it is generally more inclined to use this information for its own goals rather than leaking it to other companies. It was not possible to generalize on the type of companies that are inclined to perform profiling activities since only three traces of potential profiling activity have been detected. Out of these 3 profiling traces, 2 indicated no privacy breach and were done by the company itself, and 1 indicated a privacy breach through leakage of the customer profile to another company without data subject‘s consent. The low number of traces does not allow drawing statistical conclusions, but maybe it is fair to say that generally, a customer is able to perceive “legitimate” as well as “illegitimate” profiling activities.
3. The legitimate activities seemed to be the most frequent (11 out of 16 results). We could also observe that the ratio of companies that require agreement on a privacy policy against those who don’t is almost 1/1. Nevertheless, Bonus programs in particular always provided such a privacy policy, and were always compliant to it.
4. Privacy violations without an obvious profiling activity were also not frequent (only 3 out of 16) as far as the results could tell.
5. Due to the low number of results, it was hard to draw conclusions on correlation between profiling activity and business type (3 profiling traces from 3 different types of commercial enterprises). The common aspect between the three cases is that the registrations were done online.
6. It is important to note that none of the 6 results received from the customer loyalty programs (or bonus programs) show any trace of privacy breach or profiling activity.

7 A Scientific Look on the Experiment

As mentioned earlier, the conducted experiment falls in the category of field experiments. The purpose of the experiment was not to prove or refute a specific hypothesis, but rather to start with a broad hypothesis and gather as much perceivable results as possible within certain time and resource constraints. We think that first-hand observations of results should give clear indications towards the need for further experiments and investigations on profiling and privacy breaches in commercial institutions.

7.1 Scientific Value

It is evident that the experiment has been conducted based on a relatively broad hypothesis and a specific experimental technique. However, the experiment did not aim at deriving quantifiable scientific results. It rather aimed at obtaining empirical results that would help making observations with regard to profiling and privacy breaches in institutions that are valuable enough to incite concerned organizations to perform thorough investigations on these issues. Nevertheless, it is important to assess the methodological aspects of the experiment in terms of their scientific value, and to explain within which limits the results should be scientifically interpreted.

7.1.1 The methodology

Looking back at the methodology that has been adopted for this experiment, it is possible to interpret its conformity with scientific methodologies for field experiments.

Hypothesis

The hypothesis that we tried to verify by means of this experiment is that *profiling activities and privacy breaches on personal customer information by commercial institutions is evident and can be perceived by customers with average exposure to these institutions.* A sub part of the hypothesis states that *institutions adopting customer loyalty programs are generally more inclined to perform these kinds of activities when compared to other institutions.*

The first part of the hypothesis is easier to prove or refute than the latter part. In any case, we believe that this hypothesis should be further specified to amount for a large-scale scientific field experiment (e.g. by explicitly specifying what is considered as profiling or privacy breach by a customer). Despite its broadness, we believe that the hypothesis is suitable for the objectives of the experiment.

Variables

The variables chosen to be observed in this experiment are item-specific i.e. they reflect characteristics of each received item. We state the variables again:

- Policy compliance (True, False, N/A): a result is considered to be policy compliant if the type of the item received does not seem to violate terms and conditions agreed upon by the customer during registration to the service at the company.
- Privacy breach (True, False): A privacy breach is declared “True” if the result has been sent from a company which is different from the company at which the customer originally registered, or when there has been a violation of a privacy policy which is previously agreed on.

- Profiling trace (True, False): a profiling trace is declared when the received result is believed to be addressed to this particular customer based on a previous usage of the service by the customer. Or when the result source is different from the original company, but still is from the same company type. In both cases, the customer is believed to have been particularly targeted based on his interest.

It is clear that the characteristics of an item that we consider sufficient for declaring the value of a variable as "True" or "False" are subjectively chosen. It is also obvious that many other possible variables can be defined in the context of a scientific experiment, e.g. correlation between policy compliance and business type. Therefore, we can fairly say that the defined variables can be used for making general observations and giving hints on noticeable patterns towards the objective of the experiment.

Sampling

The sampling method used is closer to a *cluster sampling*[7] method as compared to other methods. Categories of institutions have been defined, and a few target institutions have been selected from each category. The total number of targeted institutions was 46.

Replication Aspect

The replication rate was considerably low, as has been proven by the rate of results. This is a general problem of field experiments, although we think that a scientific field experiment which is aimed at testing the same hypothesis should be able to choose a much larger sample (e.g. 200-500 institutions). This limitation in our conducted experiment has substantially affected our results despite the modesty of our objectives: it was not even possible to observe definite correlations between rates of perceived profiling activities and privacy violations and institutions type.

Randomization Aspect

We believe that the randomization method was fairly acceptable considering the objectives of the experiment. In fact, since one aspect of the experiment is to test how perceivable are profiling activities and privacy violations by normal customers, around 30 people were asked to list the types of institutions they would think of registering at. Therefore, we think that the number of clusters (17 categories of institutions) was enough, at least considering the objectives of the experiment. Nevertheless, the size of each cluster (up to 5 institutions with bonus programs) was quite low. It should be noted though that the categorization of institutions was done on a rather subjective basis, which is the similarity/distinction in the type of corresponding business.

Conditions

As mentioned at the beginning of the document, the conducted experiment falls in the category of field experiments. This means that many of the experimental conditions are hard to control, predict or even assess, a problem which is common to field experiments is general.

[7] "Cluster sampling is a sampling technique used when "natural" groupings are evident in a statistical population. It is often used in marketing research. In this technique, the total population is divided into these groups (or clusters) and a sample of the groups is selected. Then the required information is collected from the elements within each selected group. This may be done for every element in these groups or a subsample of elements may be selected within each of these groups. The technique works best when most of the variation in the population is within the groups, not between them." – Wikipedia

In particular, it is relevant to mention some of the conditions which we think have affected the variables mentioned above:

- The "terms and conditions" policy was not always available during the registration, and if it did, it was not the same for all institutions.
- Registration type was different (online, in the shop, by mailed letter, etc.) for each institution, which definitely affects the data mining and eventually the profiling activity.

Data Collection

Data collection was basically restricted to items received by post or email, and was done with a high level of care, i.e. there was no post or email that has been missed or not included in the analysis. Therefore, we think this aspect was well addressed in this field experiment. We also think that restricting the experimental data to the items received by post or email is fair enough, since this type of data is the most perceivable by the customer especially that the data is mostly of marketing nature.

Data was then categorized according to four characteristics: its form (Newsletter, brochure, etc.) its source (original company or not), its type (email or post), its company category (out of the 17 categories). The aim was to observe a correlation between the variables and those characteristics if it existed. Although other characteristics can be taken into consideration (e.g. time until item was received), we believe that this categorization was sufficient for the objectives of the experiment.

7.1.2 Scientific Interpretation

It is also important, in our opinion, to give a look into the way the results were interpreted in order to draw the conclusions and make the observations mentioned in 6.3. Two main aspects are worth considering here.

Observable evidence

The experiment undoubtedly led to observable evidence of privacy breach and profiling activities. Despite the fact that the value (true/false) declared for a variable (e.g. profiling) depended on subjectively chosen characteristics (similarity between bought item and perceived marketing), we can say that at least from the perspective of a customer, some data could definitely be considered as observable evidence. Therefore, the data was clearly of empirical nature since it was directly observed by the customer.

Validity of Interpretation Approach

We believe that the reasoning underlying the decision regarding a certain variable value was logical enough. For example, at least from the perspective of a customer, if the terms of the privacy policy he agreed on during registration are clear, it is not hard to figure out if the received marketing data is compliant with the terms of the policy. Therefore, the values assigned to the variables were accurate to a certain extent.

Interpretation of the obtained values was given in the form of observations indicating the constraint of low rate of results. The results were categorized according to "Legitimate[8]", "No Profiling, Privacy breach", "Profiling, No Privacy breach", "Profiling and Privacy breach".

[8] The compliance/violation of results to e.g. privacy policies was not assessed based on legal background or competence, but on the common sense of the customer.

This categorization was simply another form of representing the values of the variables corresponding to the experiment. Therefore, if the values given to the variables are trusted for their accuracy, this categorization can be regarded as a legitimate approach for interpreting the results.

The interpretation from which the observations in 6.3 were made was cautious enough, in the sense that no statistical evaluation was done on the values obtained, but rather observations of whether a certain type (category) of results is "perceivable/non-perceivable", or "high-frequency/low-frequency" of a certain occurrence, or an indication of some "general inclination" of institutions towards a certain type of activity. We were aware that the results were not measurable, and a statistical analysis would be inadequate especially considering the small sample of institutions and low number of results.

Based on this interpretation approach, we could verify the first part of the hypothesis, i.e. the fact that both privacy violations and profiling activities are carried out on personal data by commercial institutions. The sub part of the hypothesis was refuted based on an observation that none of the profiling traces obtained was linked to a customer loyalty program.

Therefore, it is fair to say that the approach for interpreting the results was valid, and the observations done were based on rational reasoning, as long as the values given to the variables are trusted.

7.2 Next Steps

The next steps that we propose in light of the results of this experiment are two-fold: recommendations on conducting a similar experiment with a "more" scientific approach, and general techniques that can be considered for limiting the unwanted profiling activities and privacy violations by institutions on their customers' profiles.

7.2.1 Recommendations for a more Scientific Experiment

While we leave to the reader the freedom to draw his own conclusions from the experiment, it is important to note the following recommendations if a similar experiment with strict adherence to scientific methodology is considered:

Hypothesis

The hypothesis of the experiment should be further specified, for example by more accurately defining what is perceived as privacy breach or profiling activity by the customer.

Variables

The variables of the experiment are very critical, and should be defined more accurately. We think that the choice of the variables was adequate and can be used in a scientific experiment.

Sampling

The sample size should be much bigger (over 4 times) in order to observe patterns and perform statistical evaluation. The sampling method (cluster sampling) might be the right one. Moreover, it is recommended to allocate a longer period for the experiment.

Interpretation

As profiling activities and privacy breaches are widely admitted to exist, the rate at which those activities or violations are done, and the type of institutions that are more inclined to do

them are the most critical type of results. Therefore, statistical evaluation is necessary for this experiment, and should therefore be conducted on a much larger set of data.

7.2.2 Solutions for privacy in business processes

Many solutions are under consideration for privacy in business processes. However, it is important to note that these solutions can be regarded as organizational solutions, and technical solutions.

In the following deliverable D14.8 on “Privacy in Business Processes”, a number of solutions from both categories are proposed.

The organizational approaches are mainly:

- Application standards for security and IT Service Management, and
- Data Protection Management Systems.

The technical approaches under considerations are mainly:

- Data Track
- Secure delegation of rights DREISAM
- Secure Logging and Audit
- Trusted Virtual Domains.

These approaches and techniques are thoroughly discussed in D14.8.

8 Conclusion

In this deliverable, we aimed at complementing the existing FIDIS studies on profiling and privacy compliance, with an experimental empirical study that would endorse some of the conclusions already drawn. For this purpose, we designed a framework and a methodology for collecting evidences of privacy breaches and profiling in business processes.

The experiment pivoted around the idea of name-watermarking and targeted institutions that typically allow registration of customers by providing personal data. The goal was to register at a variety of business institutions and wait until post and email letters are received from these institutions in order to derive conclusions of privacy breaches and profiling activities that are typically carried out by collecting profiles of customers and leaking them to partner businesses. The name-watermarking method aimed at tracing the process of leaking profiles to other companies in order to obtain profiling traces that can be used as a legal proof of misuse of private information.

Some conclusions can be drawn from the experiment. First, the effectiveness of the experiment can be further optimized by allowing more time for the post and emails to arrive, and by targeting more institutions.

As for the results, it could be concluded that it is hard for a normal customer who usually registers at few of these companies to perceive a considerable number of profiling or privacy breaching incidents. Even when spam advertisements are sent to post or email, many of these seem to comply with a privacy policy that the customer has already agreed on during registration. This type of policy typically allows the corresponding companies to use the personal data of the customer for their own purposes which might be different from the original intent behind registration which is providing a certain service. Therefore, these companies are legally protected against any measures that might be adopted for privacy compliance of companies with customer data.

Another conclusion was that many companies avoid proposing a privacy policy for the customer upon registration, which leaves them the space of “misusing” the personal data for their own purposes, without actually breaching any privacy policy.

Although many privacy breaches and misuse of personal data as well as profiling might intuitively be expected, the results were still pretty limited in terms of quantity. Therefore, as a general conclusion, it is necessary to note that it is relatively hard to collect substantial proofs in large quantities from corresponding companies to be used as a legal basis for accusation. The reason is that the companies seem to adhere to compliance requirements either by 1) making the customer agree a-priory on a privacy policy (even if sometimes vague enough to allow a company to use his personal data for its own purposes) or 2) by not having an agreement on a privacy policy at the first place, which can give a company some freedom in carrying out such activities if it wishes to, at least from a legal perspective.

In any case, as far as the results of the experiment could tell, customer loyalty programs (or bonus programs) are generally not inclined to violate privacy of their customer’s data or carry out profiling activities on them, at least not more than other type of companies. Customers are therefore generally unable to perceive such indication, which contradicts our original hypothesis. A larger scale and amount of data must be collected if sound collected in order to draw sound conclusions, e.g., through a web community that collects data over a longer period of time. It should be noted though that the experiment has been conducted in Germany and is not representative to other countries in Europe which might have different legislations.

Nevertheless, an assessment of the scientific value of the conducted field experiment was given, which led to the conclusion that many methodological aspects lack accuracy or adequacy. In particular, if a scientific experiment needs to be conducted towards the same objectives, the hypothesis should be more specific, the variables more accurate, and the sample size much bigger in order to be able to conduct statistical analysis on the obtained data. However, the interpretation approach that led to our observations was fairly acceptable. Next steps were proposed as recommendations for an experiment of "more" scientific nature, and techniques to combat profiling activities and privacy violations, such as organizational techniques and technical solutions.

9 Bibliography

[CIPP1] Canadian Internet Policy and Public Interest Clinic (CIPPIC), "Compliance With Canadian Data Protection Laws: Are Retailers Measuring Up?", http://www.cippic.ca/documents/bulletins/compliance_report_06-07-06_(color)_(cover-english).pdf, 2006

[CIPP2] Canadian Internet Policy and Public Interest Clinic (CIPPIC), "How Detailed Information About You Gets Into The Hands Of Organizations With Whom You Have No Relationship", http://idtrail.org/files/ExecSum_DB.pdf, 2006

[Del. 7.2] FIDIS (2005) "Del 7.2: Descriptive analysis and inventory of profiling practices", WP7, FIDIS Project.

[Del 14.3] FIDIS (2008) "Del 14.3: Descriptive analysis and inventory of profiling practices", WP14, FIDIS Project.

[ICPP] Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein, Germany (ICPP): Meike Kamp, Barbara Körffer, Martin Meints. "Profiling of Customers and Consumers - Customer Loyalty Programmes and Scoring Practices" (Book Chapter in "Profiling the European Citizen"). Springer Netherlands, May 2008.

[Hild08] Profiling the European Citizen, *Cross-Disciplinary Perspectives*, Hildebrandt, Mireille; Gutwirth, Serge (Eds.), 2008

[Hild2] Defining Profiling: A New Type of Knowledge? *in "Profiling the European Citizen"*, Hildebrandt, Mireille, 2008

[Dunn, 2005] C. Dunn. "Loyalty Programs and Privacy Issues: Do You Need to Worry About Providing Personal Information?" *Disney familty parenting*, 2005