FIDIS

Future of Identity in the Information Society

Title:	D 7.16: Profiling in Financial Institutions
Author:	WP7
Editors:	Bart Custers (TILT)
Reviewers:	Els Soenens (VUB), Ruth Halperin (LSE)
Identifier:	D 7.16 Profiling in Financial Institutions
Type:	Final report
Version:	1.0
Date:	Monday, 29 June 2009
Status:	Final report
Class:	Final report
File:	Final report 7.16 v1.0.doc

Summary

Financial institutions have both business incentives and legal obligations to create risk profiles of their clients. Implementing profiling policies, however, raises several problems and does not seem to be effective and efficient for the purposes intended, such as risk management or tracking fraud, money laundering and terrorist funding.

In this deliverable, it is investigated how risk profiling strategies are implemented in practical environments in financial institutions and which practical problems this may cause. The focus is on both legal and informational aspects. From a legal perspective it is investigated whether translation of legislation into practical policies and procedures has the intended effects. From an informational perspective it is investigated whether the systems and analyses used for risk assessments have the intended effects.

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

Copyright Notice:

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)*
0.01	04.02.2009	• First version with contributions of all participants
0.02	06.02.2009	• 5.3 added
0.03	10.02.2009	• Additions Bart Custers
0.04	25.02.2009	• Section 3.3 on systemic risks added, conclusions on IP law added, key question added
0.05	02.03.2009	• Section 3.1 and 3.5 added, minor revisions
0.06	20.04.2009	• Conclusions added, section 3.4 (VAT fraud) removed. [3.5 renumbered into 3.4], glossary added.
0.07	19.05.2009	• Summary added, final remarks of all participants included
1.0	29.06.2009	• Revision after internal review

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 Summary	Bart Custers (TILT)
2 Introduction	Bart Custers (2.1, 2.2), Martin Meints (ICPP) (2.2, 2.3), Niels van Dijk (VUB) (2.3.2)
3 Profiling in the Financial Sector	Martin Meints (3, 3.2.2, 3.4), Martin Meints and Harald Zwingelberg (ICPP) (3.1), Rainer Böhme (TUD) (3.2, 3.2.1, 3.2.3, 3.3)
4 Problem Statement	Bart Custers (4.1, 4.2, 4.3), Martin Meints (4.4)
5 Legal Regimes	Bart Custers (5.1), Niels van Dijk (5.2), Harald Zwingelberg (5.3)
6 Conclusions	Bart Custers

Table of Contents

1 Executive Summary

Financial institutions are increasingly profiling their clients for different reasons. The first reason is for risk management (i.e., assessing risk in order to prevent costs and payments from exceeding revenues). The second reason is for compliance (i.e., adhering to laws and regulations). These regulations mainly include taking measures against fraud, money laundering and terrorist funding. In recent years, strict legal measures (such as Sarbanes-Oxley and Basel II) were introduced to increase financial stability.

Implementing profiling policies, however, raises several problems and does not seem to be effective and efficient, both from a business perspective and from a compliance perspective. From a business perspective, risk assessments have been proven ineffective in the current financial crisis, where unacceptable risks were taken. From a compliance perspective, risk assessments are also ineffective, as it is not too difficult for criminals and terrorists to avoid being noticed during these types of screening. Hence, the key question of this research is: how do financial institutions implement risk profiling strategies from a legal and an informational perspective and with what effects?

This question is answered by investigating areas of application of profiling in the financial sector, such as fraud prevention in payment transactions, credit reporting and credit scoring, prevention of systemic risks and anti-money-laundering and the prevention of terror-financing. These cases, particularly the latter case, are described and analysed from a technological perspective (regarding the methods of implementation, the practical problems of implementation and the general advantages and disadvantages of profiling) and from a legal perspective (regarding the applicability of different legal regimes and the protection these legal regimes may offer for those involved). This research is based on available literature and the working experience of the authors in the financial sector in their respective countries.

From an informational perspective, the risk profiling strategies involve gathering large amounts of data by financial institutions. These data, though not always correct and useful, are usually analysed in order to find risk profiles. Although these risk profiles may be lacking reliability, they are applied to take measures against high risk clients. These high risk clients may be put under close scrutiny, rejected financial services, blacklisted, etc. Clients often have little means of redress as transparency regarding profiling and its implications is lacking.

From a legal perspective, risk profiling strategies usually concern avoiding fines for non-compliance, although some compliance measures may go hand in hand with business goals. For financial institutions it is not always clear how to properly implement legislation. Therefore, they often (try to) discuss strategies with supervisory authorities. The supervisory authorities, however, are often reluctant to provide advice on this, as it may be difficult to apply enforcement afterwards.

The intended effects of risk profiling strategies are to deal with risk properly (including creating financial stability and avoid unacceptable risk) and to deal with crime (including fraud, money laundering and terrorism financing). In general, it can be said that these effects are not achieved with the current profiling strategies that financial institutions use. As mentioned above, the financial crisis has shown that unacceptable risks were taken.

Furthermore, it is not too difficult for criminals and terrorists to avoid being noticed in profiling procedures.

Informational recommendations are:

- Carefully assess which data to collect and use for analysis. Collecting less data may ensure a more targeted approach and, at the same time, causes less security and privacy issues.
- Regularly update profiles, as risks constantly change.
- Focus on transparency, particularly regarding identifying key decision makers and countries with favourable tax climates and strict banking secrets.
- Do not use the standardised approach of profiling large numbers of customers. Instead, use expert teams for in-depth investigations on suspects identified by the profiling processes. Combine technology with human intuition.
- Create possibilities of redress for cases in which errors occur.

Legal recommendations are:

- Lawmakers should focus on the intended effects of laws and recommendations, rather then on prescribing *how* these goals should be achieved.
- Create more uniformity in the legal regimes protecting the stakeholders, ensuring that they do not conflict. Particularly a clearer balance between personal data protection and intellectual rights is needed. Obviously, the previous recommendation may also help in this.
- Financial institutions should have clear and open policies on how they collect and process personal data, including statements on which data they collect and for what purposes. Focus on the gap between policy and practice is also important and requires stronger enforcement.
- Supervisory authorities (particularly national data protection officers and financial watchdogs) should particularly exercise stronger enforcement regarding transparency.

2 Introduction

2.1 Scope

This report deals with profiling in financial institutions, such as banks, insurance companies and credit card companies. Profiling is very common in the financial sector and, when looking from a risk management perspective, it may even be suggested that it is their *core business*, i.e., their essential activity. For instance, banks have to assess risks of non-repayment of loans and mortgages and insurance companies have to assess risks in order to prevent that their payments (where risk occurs) do not exceed their revenues (premiums paid).

Profiling technologies have become more advanced in recent years. Large amounts of data have been collected and stored in databases and new technologies have been developed to do profiling and risk assessment. Apart from these developments, profiling and risk assessment is nowadays also playing a role in *compliance*, i.e., adhering to laws and regulations. In order to track fraud, money laundering and terrorist funding, financial institutions have a legal obligation to create risk profiles of their clients. Banks that want to do business in the United States have to implement a worldwide Know Your Customer (KYC) program,[4] partially based on the Patriot Act.[5] Most European countries have implemented similar legislation after the terrorists' attacks in London and Madrid.

However, the implementation of such profiling policies raises several problems and does not seem to be effective and efficient. From a business perspective, risk assessments have been proven ineffective in the current financial crisis, where unacceptable risks were taken.[6] From a compliance perspective, risk assessments are also ineffective, as it is not too difficult for criminals and terrorists to avoid being noticed during these types of screening.[7]

In this deliverable, it will be investigated how risk profiling strategies are implemented in practical environments in financial institutions and which practical problems this may cause. The focus will be on both legal and informational aspects. From a legal perspective it will be investigated whether translation of legislation into practical policies and procedures has the intended effects. From an informational perspective it will be investigated whether the systems and analyses used for risk assessments have the intended effects. Hence, the key question of this report is:

How do financial institutions implement risk profiling strategies from a legal and an informational perspective and with what effects?

The research that was done to answer this question is based on available literature and the working experience of the authors in the financial sector in their respective countries. In chapter 2, the terminology and the profiling technologies used in this report are discussed. In chapter 3, areas of application are discussed. These cases focus on business (credit reporting

[4] A KYC program involves identification and risk assessment of all clients of a financial institution.
[5] Patriot Act, 2006, see <http://www.epic.org/privacy/terrorism/hr3162.html>
[6] See, for instance, Fratianni and Marchionne (2009). Even before the financial crisis there were indications of poor risk assessments, see, for instance, Wyatt (2002), Kliger and Sarig (2000), Galil (2003).
[7] See, for instance, Birrer (2006), Jonas and Harper (2006), National Research Council (2008).

and credit scoring), on compliance (tax evasion, anti-money-laundering and prevention of terror-financing), or both (fraud prevention, systemic risks). In chapter 4, the implementation of profiling strategies is discussed, including practical problems that arise from implementation and advantages and disadvantages of profiling. In chapter 5, a legal approach is taken. The framework of privacy and data protection is discussed, as is the framework of intellectual rights. It is investigated to which extent these frameworks may apply to profiling strategies. Chapter 6 provides conclusions and answers the research question.

2.2 Terminology

This report deals with several legal and technological concepts that are sometimes used in different ways. Therefore, we start with describing how these concepts are used in this report. Basic terminology in the context of profiling already was developed and discussed by Vedder (1999), Custers (2004) and Hildebrandt and Backhouse (2005). In the context of profiling in the financial sector the following terms may play a role:

Profiling

a. the process of constructing profiles
   i.e., constructing profiles (correlated data), that identify and represent either a specific person or a group/category/cluster.
b. the process of ascribing profiles to individuals or groups
   i.e., 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.

Profiles

A profile is a property or a set of properties of a specific person or a group/category/cluster. Generally profiles are distinguished in *personalised profiles* (also referred to as individual profiles, personal profiles, customer profiles, specific profiles, etc.) and *group profiles* (also referred to as aggregated profiles, abstract profiles, etc.)

Personalised Profile

A personalised profile is a set of correlated data that identifies and represents a single person.

Group Profile

A group profile is a set of correlated data that identifies a group, and/or when applied identifies a person as a member of a group.

Not all group profiles are alike. There are different ways to distinguish types of group profiles. An important distinction concerns the way information is used in forming group profiles. Another distinction is that between distributive and non-distributive profiles, which is particularly important for the reliability of (applying) group profiles.

Associations are made when comparable events are connected with each other. For instance, a person who is overweight may have an 80 percent chance of having high blood pressure as well. Association rules generally have a so-called if-then structure. If a person has characteristic X, then she has a 70 percent chance of also having characteristic Y.

Sequences are connected successive events. For instance, patients may have a 40 percent chance of getting the flu, after hospitalisation. The difference between associations and sequences concerns time: associations are made at the same time, while sequences involve a certain time to have elapsed between events. Thus, sequences usually have the if-then structure that associations have as well, though sequences contain a time adjunct. When the time elapsed between the two events is very short, the distinction between a sequence and an association disappears.

Classification is the examination of known groups to determine which characteristics may be used to identify or predict group membership; for example, classification by nationality. Particular properties, such as language or ethnic background, may be used to identify or predict group membership. Thus, a person speaking Icelandic is likely to have Icelandic nationality. A person of Albanian ethnic origin may have Albanian nationality, but this prediction is less reliable since there are also many Albanians with Yugoslav or Macedonian nationality.

Clustering is used to discover different groups within the data. Thus, a particular characteristic is first chosen, and using this characteristic, it is then investigated whether different groups may be identified. For instance, taking the characteristic "age", it might be discovered that many middle-aged people live in the countryside, while young and elderly people mainly live in cities. If this were the case, then age may be clustered around young, middle-aged, and elderly people, each having its particular characteristics.

Classification and clustering may be used to identify groups. The difference between classification and clustering is that classification uses known groups to find characteristics, while clustering uses known characteristics to find groups.

Predictions are time extrapolations of parameters. Extrapolations are usually done on the basis of regression functions, but the use of more complex methods, using dynamic approaches, is also possible. Various parameters may be predicted, but in general, the further ahead, the less reliable the results are. For instance, making predictions about next month's weather is useless because it is simply too far ahead to make sense. The results of predictions are also less reliable when the data on which the prediction is based do not go back far in time. For instance, it is generally held that predicting next week's Wall Street figures on the basis of last week's figures is less reliable than doing so on the basis of last year's figures. Predictions make it possible to anticipate events, and are thus similar to sequences. The difference between predictions and sequences is that predictions concern parameters, while sequences concern events. However, the predicted values of certain parameters may make it possible to find sequences.

Another important way of distinguishing different types of group profiles concerns the distributivity of properties.

Distributive Profiles

In a distributive profile, the properties of the profile are valid for each individual member of a group.

Non-distributive Profiles
In a non-distributive profile, the properties of the profile are valid for individuals as members of that group, though not for those individuals as such.[8]

The following example may help to understand the concept of distributivity. Take for instance the Smith family. This group consists of John and Mary and their children Bob, William and Susan and Donna. All six family members are blond. This is a distributive profile, as the profile is also valid for the individual family members (i.e., John is blond, Mary is blond, etc.). Half of the Smith family wears glasses: John, Bob and Susan wear glasses, but the other family members do not. The profile "members of the Smith family have 50 % likeliness of wearing glasses" is non-distributive. This profile is true for each member of the group, but not for individuals as such (e.g. John as an individual wears glasses and Donna does not).

Note that whether or not it is possible to determine the distributivity of a group profile may depend on whether the perspective is internal, i.e. from inside the group, or external, i.e., from outside the group. For instance, in a group in which 75 percent of the people have blond hair, an individual group member with black hair may claim the profile is non-distributive. Nevertheless, an outsider will ascribe a 75 percent chance of being blond to each (random) group member. Although it is possible to falsify a distributive profile by using only one appropriate (falsifying) example, verifying a distributive profile requires the examination of each individual group member. Still, in the case of future characteristics (for instance, life expectancy), examination is possible only after the profile has "expired", i.e., when the claimed future characteristic has become present or should have become present by that time. Non-distributivity is closely connected with the reliability of the use of group profiles. Note that personalised profiles are always distributive, as the group counts only one member.[9]

Risk profiles
In financial institutions, profiles are often referred to as risk profiles. Risk profiles are profiles (either personalised profiles or group profiles) that indicate a particular risk, such as creditworthiness, risk on money-laundering or risk on fraud.[10]

Financial Institutions
Financial institutions in this report are institutions that provide financial services for its clients or members. This includes banks, insurance companies, accountancy firms, pension funds, investment funds and credit card firms.

[8] Vedder, A.H. (1996b) Privacy en woorden die tekort schieten, In: *Privacy in het informatietijdperk*, S. Nouwt and W. Voermans (eds.) Den Haag: SDU Uitgevers.
[9] Note that personalisation of services can also be based on non-distributive profiles. In fact, this is often the case when organisations are working with incomplete data or when they prefer working with aggregated data.
[10] Note that this definition depends on what is considered a risk.

2.3 Profiling Methodologies

2.3.1 Goals and methods of profiling

The different types of group profiles discussed in the previous section are the result of different profiling methodologies. Profiling can be carried out based on different concepts using different methods. Relevant in this concept are (Schweizer 1999: 57-62; Oberlé 2000):

1. Detection of association rules (example: with X% of likeliness some buying beer also will buy potato crisps)
2. Detection of sequencing rules (example: in case event A happens with X% of likeliness two weeks later event B will happen)
3. Classification (analysed data sets are assigned to one of a number of pre-defined classes; this also is referred to as directed data analysis)
4. Clustering (analysed data is clustered based on similarity detected in the analysis; non-directed data analysis)
5. Predictions (time extrapolations, i.e., predicting X in the sequence 2, 4, 6, 8, X)

The main methods for constructing profiles are regression, clustering and classification. These methods are illustrated in Figure 2.1.

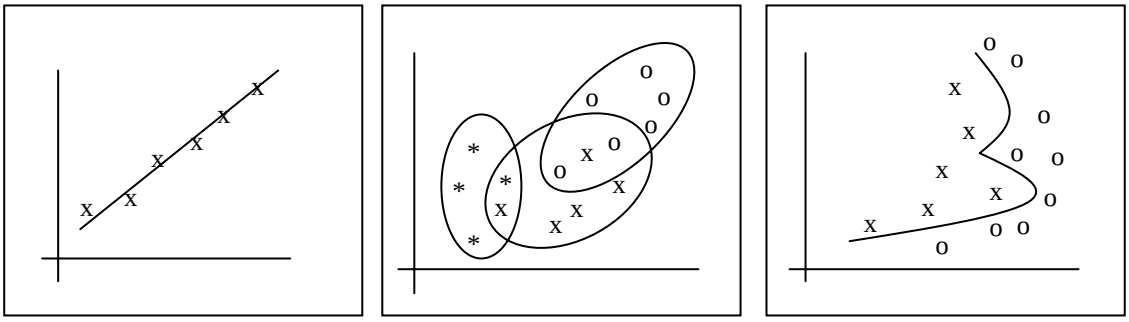

Figure 2.1: *Examples of different types of discovery algorithms: Regression (A), clustering (B), and classification (C).*

The purpose of *regression* is to describe data using a function. In Figure 2.1A, the data is represented by a linear function. A typical example of a linear relation is the relation between shoe size and tallness: taller persons have, in general, larger feet. And the taller the person, the larger his or her feet will be. *Clustering* is used to describe data by forming groups with similar properties. In Figure 2.1B, three different groups are identified, marked by stars (*), open dots (o) and crosses (x). After identification, descriptions of these groups may be found, indicated by the ellipses drawn. Note that the groups may overlap. *Classification* is used to map data into several predefined classes. In Figure 2.1C, a predefined class boundary is drawn (a non-linear curve), creating two classes (one to the left of the curve and one to the right of the curve). After the class boundary is defined, each data subject is classified into one of the two classes. Once it is clear to which class each data subject belongs, it is possible to attach labels, which is done by attaching crosses (x) and open dots (o).[11]

[11] Note that overlap is not possible in the case of classification.

In the literature, many other types of algorithms are mentioned, but most of them may be accounted for by the three types mentioned here.[12] These three types of data mining may be relevant to group profiling. There is a plethora of algorithms that can be used to do regression, clustering and classification. These algorithms, such as decision trees, including classification and regression trees (CART), chi-squared automated detection (CAID), logistic regression,[13] genetic algorithms, Bayesian networks and neural networks. These algorithms will not be discussed here. For an overview, see Fayyad et al (1996).

2.3.2 Steps of the profiling process

The profiling process is usually associated with the process called KDD, i.e. *Knowledge Discovery in Databases*. In this context, the term data mining is often used. *Data mining* is a technique used to discover hidden information in very large databases. Data mining is a step in a process called (KDD).[14] Fayyad et al. (1996) define Knowledge Discovery in Databases as the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. This process consists of five successive steps, as is shown in Figure 2.2. In this section, it is briefly explained how profiling using data mining takes place.

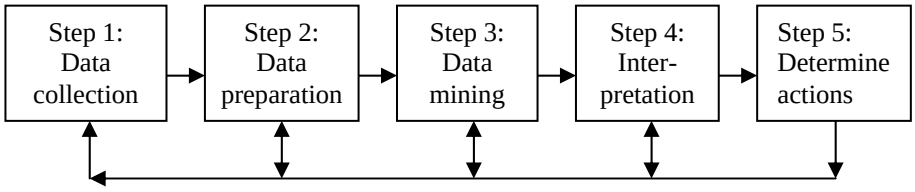

Figure 2.2: *Steps in the KDD process*

- Data Collection: The target dataset or *database* for analysis is formed by selecting the relevant *data*
- Data Preparation: The data are pre-processed for removing noise and reducing complexity
- Data Mining: analysis of the data with the *algorithm* or *heuristics* chosen to suit the data, model and goals
- Interpretation: the mined *profiles* are evaluated on their usefulness by the analysts
- Determine actions: identification of a user as a member of a profile and institutional decisions and actions upon profile application

2.3.3 Good practice frameworks for the profiling process

For the process of profiling, especially for good practice in roles and tasks, frameworks are available. Examples already were discussed in (Hildebrandt, Backhouse 2005). Especially important are:

[12] E.g., *decision trees* may be further divided into classification trees and regression trees.
[13] See for details on logistic regression http://www.metaanalyse.de/material/re020610.pdf
[14] The KDD process is sometimes referred to as data mining, which may be confusing. This report refers to data mining only as a step in the KDD process to avoid such confusion.

- Cross-Industry Standard Process for Data Mining (CRISP-DM), currently version 1.0[15]; currently work is carried out for the development of version 2.0 (also see Hildebrandt, Backhouse 2005: 20-22)
- Semiotic Model of Knowledge Discovery in Databases (KDD), see (also see Hildebrandt, Backhouse 2005: 22-26 and literature cited therein)

[15] See www.crisp-dm.org

3 Profiling in the Financial Sector - Areas of Application

This report deals with profiling in financial institutions. In the financial sector, typically banks and insurance companies carry out the profiling, while different interest groups or stakeholders are involved. These stakeholders are:

- Financial institutions, often for their own financial benefit or risk management
- Society, represented by the state,
- Individuals in their roles as customers of the financial institutions (mainly direct interest) and citizens/members of the society (mainly indirect interest).

In this chapter, several cases are described from the perspectives of each of these stakeholders. This approach was chosen since the interests of these stakeholders may be conflicting, even though in many cases it is up to the banks to carry out the profiling and to implement balanced solutions taking care of the different interests. Table 3.1 gives an overview on the applications of profiling in the financial sector described and analysed in this report (left column). The direct interests of the stakeholders, for which most awareness seems to exist, are described in the table.

		Stakeholders		
		Financial Institutions (FI)	**Individuals**	**Society**
Cases/Areas of application	**Fraud Prevention** (section 3.1)	Protection of money (liability)	Protection of money (individual as customer)	Protection of privacy as a social right balancing (a) the need to profile to find unlawful intrusions of private life/personal information by criminals and (b) the aim to respect personal rights of citizens by state institutions)
	Credit Reporting / Credit Scoring (section 3.2)	Risk management	Exclusion to loans, mortgages, insurance	Protection of the stability of the financial sector, see Basel II
	Prevention of Systemic Risks (section 3.3)	Preventing Bankruptcy		Protection of the stability of the financial sector
	Anti-Money-Laundering (section 3.4)	Preventing fines		Fighting organised crime and terror organisations

Table 3.1: Direct interests of the main stakeholders in the cases described in this chapter.

It is important to note that these perspectives are partly subjective and simplified, as they do not take into account all roles of the stakeholders and interconnections between the described targets and methods of scoring used. These perspectives were chosen to as they cover the viewpoints of the most important stakeholders and typically illustrates their interests.

Individuals may not value and as a result do not put forward their *indirect interests* (in this context their interests as members of the society in the stability of the financial sector or the prevention of crime) in the same way as they value and express their *direct interests* of secure savings and easy access to credit money. Already these two direct interests may be conflicting, as we see in the current crisis in the financial sector. Too easily granted access to credits mainly in the USA put the security of savings into world-wide danger.

An example of the interconnections of methods is the use of credit scoring, which is optimised for the management of risks of the financial institution, but at the same time can play a role in stabilising the whole financial sector on a macro economic level (prevention of systemic risks).[16] This concept was commonly expressed before the current crisis in the financial sector since 2007, for instance in the BASEL II regulations.[17] In addition, financial institutions claim that there also is an impact of credit scoring on the economic well-being of individual customers, though this statement is not supported in the same way by consumer protection organisations.

3.1 Fraud Prevention in Payment Transactions

3.1.1 Introduction and Legal Grounds

Fraud prevention in payment transactions has many aspects, as a number of different parties with different roles and various technical solutions are involved. Fraud prevention measures include:

- Security of payment terminals and ATMs
- Securing network transport of payment transaction data
- Securing systems of banks, points of acceptance, or card organisations
- Security of physical transport e.g. of payment cards
- Technical security functions such as photos, holograms, checksums, cryptographic techniques etc. on payment cards
- Monitoring of payment transaction data by applying profiling techniques to detect potential fraud

In the context of this deliverable we will focus on monitoring practice concerning payment transaction data. As an example we will take a look at credit card based payments, as they

[16] *Systemic risk* is the risk of collapse of an entire financial system or entire market, as opposed to risk associated with any individual entity, group or component of a system. Kaufman and Scott (2003).

[17] Basel II is the second Basel Accord with recommendations on banking laws and regulations, see http://www.bis.org/publ/bcbsca.htm. These recommendations aim to provide an international standards for banking regulators when evaluating the risk management of banks. In general, banks need to hold to safeguard more capital when exposed to greater risk.

specifically rely on the monitoring of payment transaction data due to their capability for payments where the card is not present (card-not-present transactions) and as a result the non-applicability of some of the mentioned security measures.

So far the prevention of fraud in payment transactions has no specific legal grounds in EC legislation.[18] This is going to change soon. In the context of the harmonisation of European payment transactions end of January 2008 the Single Euro Payment Area (SEPA) went operative. SEPA includes European technical standards e.g. for transaction formats and common European legal grounds. Important in the context of the legal grounds is the Directive 2007/64/EC of Payment Services in the Internal Market.[19] Art. 79 of this Directive states: "Member States shall permit the processing of personal data by payment systems and payment service providers when this is necessary to safeguard the prevention, investigation and detection of payment fraud. The processing of such personal data shall be carried out in accordance with Directive 95/46/EC." The implementation of this Directive in European national legislation is mandatory by end of October 2009. Once transposed the national rule corresponding to Art. 79 of Directive 2007/64/EC will provide a specialised rule for this type of data processing and the corresponding change of purpose of the payment information for the purpose of fraud prevention.

3.1.2 Fraud Prevention Practice in Credit Card Payment

When looking into credit card based payment we typically find the following stakeholders, roles and transactions:

[18] But see the recommendation of the Basel Committee on Banking Supervision on Customer due diligence for banks which sets comparable requirements for account monitoring.
[19] Amended 13th of November 2007, see http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:319:0001:01:EN:HTML

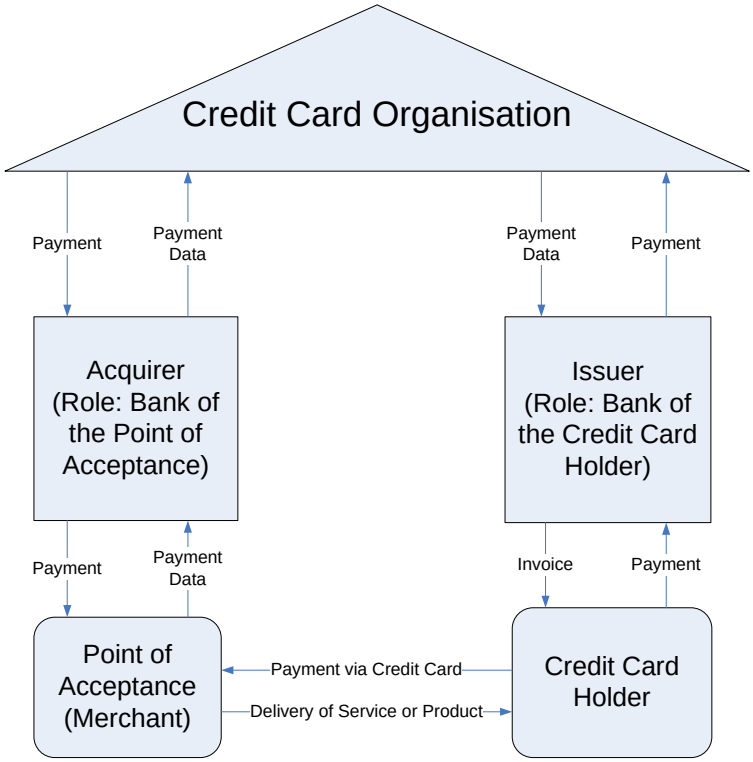

Figure 1: Stakeholders, roles and transactions in credit card based payments[20]

Monitoring of payment transactions concerning a credit card is carried out by the issuer, as he has access to all payment transactions concerning cards issued by him. But acquirers also offer this service to their clients.[21]

Details on the profiling process are not publically accessible, obviously due to reasons of secrecy; issuers obviously fear that methods used get less effective or bypassed once known by fraudsters. As a result the categorisation of the profiling applied is not possible. However, basic information on criteria used is available.[21] These criteria include:

- order, buyer address data,
- location data (discrepancy between claimed and real location),
- black list of blocked or abused credit cards and other data,
- payment behaviour of the card holder concerning products and online shops,
- stability of certain data such as IP address, card data and linked bank account via a certain period of time,
- certain "logical pattern concerning buyer and payment" and
- number of payment transactions and amount of money transferred in a certain time period.

[20] Scheme based on tutorial material provided by b+s Card Service GmbH, Hamburg
[21] See for example WorldPay, http://www.worldpay.com/support/deutsch/content.php?page=fraud&sub=tools&subsub=riskmanagement

From the information that issuers,[22] acquirers[21] and product vendors for profiling solutions[23] publish, the following generic process for fraud prevention concerning credit card based payments can be derived:

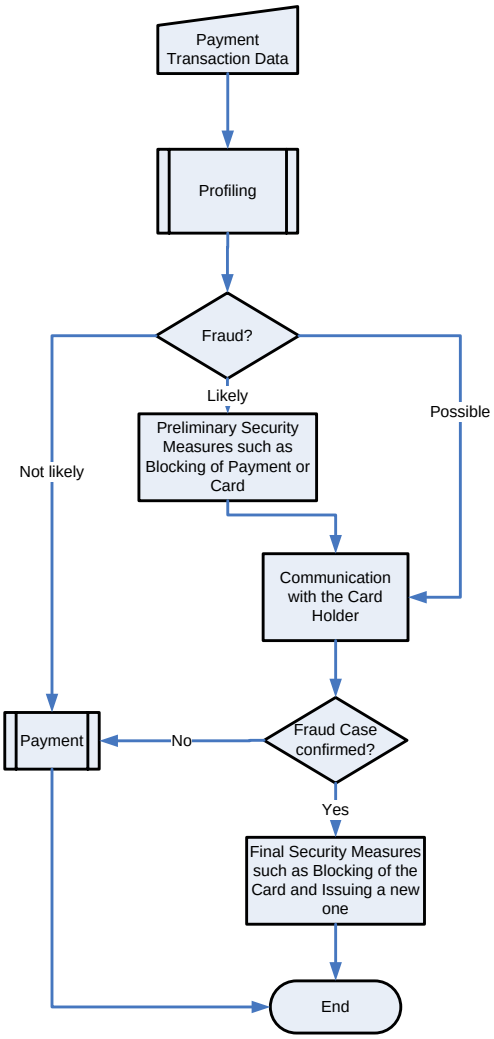

Figure 2: Prototypic fraud prevention process for credit card based payment

3.1.3 Data Protection Aspects

The legal basis for the data processing for the purpose of fraud prevention lies within the national rules corresponding to Art. 79 of Directive 2007/64/EC. Prior to the transposition of the directive data processing may be based on Art. 7 (c) of the Data Protection Directive. That article allows data processing that is necessary to fulfil the duty to report suspicious transactions to the appropriate authorities, processing the data is therefore necessary to comply with a legal obligation. On-going monitoring of accounts to understand the normal

[22] See e.g. http://www.worldpay.com/support/deutsch/content.php?page=fraud&sub=risk
[23] See e.g. http://swbplus.bsz-bw.de/bsz277201586inh.htm

and reasonable account activity of the customers is necessary to identify transactions that fall outside the regular pattern (Basel Committee, 2001: 13-14). At least in Germany the rules set forth by the Basel Committee were declared mandatory for German banks and therefore constitute a legal obligation in the sense of Art. 7 (c) of the Data Protection Directive. The processing of the customers payment data can furthermore be based on Art. 7 (f) of the Data Protection Directive as the prevention of fraud is in the interest of the financial institution to prevent customers claims.[24] Overriding interests of the data subject are unlikely as fraud prevention also takes place in the interest of the data subjects. Even if the customers do not lose any rights and claims against the credit card company they still have to deal with the trouble and possibly lack of evidence for not sharing any private tokens or PINs.

In regard to data protection the above described process of credit card monitoring for the purpose of fraud prevention raises several issues:

- transparency, in particular information of the data subject and
- automated individual decision making.

Transparency and Information of the data subject

Articles 10 and 11 of the Data Protection Directive require that the controller must inform the data subject with regard to the identity of the controller, purposes of the processing, possible recipients and the right of access. However, in practice it is to assume that most member states applied the exception to Articles 10 and 11 provided in Art. 13 of the Data Protection Directive in general or at least for credit card and bank account monitoring concerning purposes of public security and prevention and detection of criminal offences.[25] At least after the transformation of the requirements of the Directive on Payment Services in the Internal Market (2007/64/EC) it can be assumed that the member states have provided for corresponding provisions. We analysed a selection of European credit card contracts. Even though the privacy policies had high standards with regard to information on data processing concerning credit scoring and other issues, information on the legally required general monitoring of transactions for the prevention of money laundering or fraud had not been given.

Automated individual decisions

Another requirement in respect to data protection regards possible automated individual decisions made by the monitoring system. According to Art. 15 of the Data Protection Directive member states "shall grant the right to every person not to be subject to a decision which produces legal effects concerning him or significantly affects him and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc."

The term "data intended to evaluate certain personal aspects" refers to a profile on the personality consisting of several pieces of information regarding an individual (Ehmann, Helfrich, 1999: Art. 15 note 12-14). The cumulated payment data and its analysis to determine a regular pattern of the customer's payment behaviour constitutes such a profile.

[24] Cf. for the applicability in whistleblower schemes: Art. 29 Working Party, Working Paper 117, Opinion 1/2006 p. 7-8. Download: http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2006/wp117_en.pdf

[25] An example for a general application of Art. 15 is the German Data Protection Act (BDSG) according to which it is not necessary to provide the information, when the data processing is explicitly required by law. See §§ 19a Sec. 2 Nr. 3 and § 33 Sec. 3 Nr. 4 BDSG.

Furthermore the monitoring of the payments usually occurs fully. And finally a rejection of a payment may significantly affect the cardholder, in particular when the card holder is travelling abroad and relying most on the availability of the credit card. Thus Art. 15 of the Directive is in principle applicable to measures of credit card companies' monitoring systems.

However, this does not count for automated decisions that are not based on a profile but previously defined conditions, e.g. denying payment as the daily or overall maximum balance of the card is exceeded. This does not constitute an automated decision in the sense of Art. 15 of the Data Protection Directive.

Even if the system finds a payment to be suspicious of fraudulent abuse, actions with minor effects on the cardholder may be initiated by an automated system as only such measures are forbidden that significantly affect the data subject. So for example it would be feasible that the system demands further identification of the credit card holder prior to approving a transaction.

This leaves cases that require for a definite denial of the payment or even a revocation of the card. Such actions may not be based on an automated decision. But it is possible that the system notifies a human operator, who then takes further measures based on the facts of the individual case. Further measures may include a direct contact with the card holder to further verify the situation. If a contact and thus verification is not possible, the operator may take individual decisions.

Conclusion
The monitoring of credit card use raises some questions in regard to privacy and data protection. Particularly in cases of automated decisions, a decision of a human operator to approve measures is necessary to ensure compliance with the Data Protection Directive.

3.2 Credit Reporting / Credit Scoring

In most countries, lenders routinely share information on the creditworthiness of their borrowers through credit bureaus, either owned by the lenders themselves or by a third party. In this context, the term "lenders" is quite broad and includes banks and financial services, but also retailers, suppliers (via trade credit), utilities, and telecommunication companies.

Lenders supply the bureau with data about their customers. The bureau builds a personal profile for each borrower, also taking into account information from other sources, such as public registers and courts. This profile is later shared with creditors, who use these "credit reports" to decide about credit applications. Nowadays, this information is conveyed in digital format over computer networks. The process is largely standardised and often automated.

Economic theory suggests that information sharing has threefold effect on credit markets (Jappelli and Pagano 2002):

- Primarily and most obviously, information sharing among lenders attenuates adverse selection and enables risk-adjusted pricing. Lenders grant applicants with high creditworthiness better conditions, whereas those with questionable creditworthiness

are charged higher interest rates to compensate for their risk of default or are even turned down completely.

- Similar to reputation systems, the existence of information sharing mechanisms has a disciplinary effect on borrowers. Knowing that bad behaviour will lead to sanctions in future transactions prevents moral hazard and creates incentives for borrowers to repay present debt.
- In making available crucial information equally to all market participants, information sharing mechanisms reduce the information rent of house banks and increase competition for loans. (House banks would otherwise possess private information and offer to own clients of high creditworthiness conditions just under the market price for unknown borrowers). Overall, this leads to a decrease in net interest rates.

All three mechanisms agree on the prediction that information sharing among creditors *ceteris paribus* reduces default rates, but the prediction on lending remains equivocal even in theory. On the one hand, lending is boosted by better credit conditions to high-quality borrowers. On the other hand, lending may be negatively affected by the suppression of information rents that allowed banks to partly offset losses from lower-quality loans. These loans turn uneconomical in the presence of intense competition due to information sharing and will not be granted anymore.

3.2.1 Credit Reporting

This section outlines the structure of credit reporting systems in Europe. In general, the industry is characterised by heterogeneity across countries (Jentzsch 2007), primarily reflecting the historic development of credit markets and differences in the legal and regulatory regimes. Credit reporting mechanisms can be broadly classified along the dimensions ownership, type of information shared, and credit market segment.

Ownership structure
A distinction can be made between private credit bureaus and public credit registers. Among private bureaus, many are founded and still owned by a consortium of lenders. In this case, the bureau's objective is not primarily to generate profit on its own. Its raison d`être is rather justified by the saving realised on the lenders' side due to reduced credit risk and moral hazard. Since credit reports are information goods, the most efficient market structure is a single information broker that interacts with all lenders on the market. Only recently, profit-oriented credit bureaus entered the market, mainly in attempts to grow beyond national borders in anticipation of an increasingly harmonised European or even global credit market.

State-controlled information sharing mechanisms appear in different forms. Most countries have established special registers for real estate collateral. These registers list seniority of rights as well as bankruptcy information to warn creditors and potential new lenders. In some countries, additional public credit registers (PCRs) exist. PCRs can operate very similar to credit bureaus and various forms of organisation can be observed. In the large majority of countries, PCRs are managed by the central bank, in some cases by the financial supervision authority, and there exist also a few other models, e.g., Finland has outsourced the operation of its PCR to a private company.

While credit bureaus and PCRs operate in a similar fashion (and fulfil the same functions), the main difference between them is that reporting to the PCR is often mandatory by law. As a result, only PCRs achieve universal coverage for their specific type of information (also defined by law). In some countries (e.g., Germany), the PCR primarily exists for the purpose of prudential supervision. As a consequence, its focus lies on large clusters of debt and therefore high reporting thresholds apply (e.g., 1.5 million euro in Germany). In countries where PCRs cannot fulfil all information demands of private lenders (due to reporting thresholds or restrictive access policies), PCRs and credit bureaus often co-exist.

Type of information shared
Information in credit reports can be broadly divided into positive ("white") and negative ("black") information. *Negative information* is usually much more sparse and typically includes statements on

- past defaults, and
- arrears.

Positive information typically allows to draw much more detailed pictures (profiles) of larger parts of the population, including statements on the data subject's

- assets and liabilities,
- debt maturity structure and guarantees,
- employment,
- income,
- family,
- relocation history, and
- past credit inquiries.

This list applies to consumer credit reports. Corresponding items are collected and exchanged for corporate borrowers. Both negative and positive information have advantages and disadvantages. Sharing only negative information appears to better reflect the data avoidance principles laid down in data protection legislation. Further, databases of negative information contain entries of fewer individuals and are less susceptible to be abused as a means of mass surveillance. However, from a creditor's point of view, the difficulty of reliable re-identification creates a security problem. Applicants have an incentive to get rid of bad reputation by cutting or blurring the link to old records (i.e., change of name and country of residence). Contrary, a convention to provide reference to extensive sets of positive information, similar to a "financial curriculum vitae", gives raise to the problem of identity theft. Swindlers may abuse the identifying information (which often contains no secret) of innocent victims to obtain credit on the basis of their bona fide reputation.

Loan size and quantity
With regard to loan size and quantity, the market can be divided into two segments. On the one hand, large corporations as borrowers usually require a complex assessment of creditworthiness. This process is difficult to standardise and considers information such as balance sheet indicators, ownership and management information, relation to parent companies and subsidiaries and the general market environment as measured by macro-economic and industry-specific factors. This is the typical business of a few specialised rating agencies with global presence. On the other hand, information sharing and aggregation for small businesses, consumer, and trade credits is very different. Credit decisions and events

occur in large numbers, thus allowing employing statistical tools, and defining standardised – to a certain degree even automated – processes. Construction and application of group profiles is pertinent in this business.

Cross-country overview

Table 3.2 provides a snapshot of the structure of credit reporting systems in EU member states as of 2005/2006. Private credit bureaus exist in all countries but France (and Luxembourg), and they all collect negative information. Positive information is collected in most countries and the few exceptions are largely due to legal restrictions. PCRs exist in about 50% of the countries and, with the exception of France, complement at least one private bureau in the same country. It is remarkable that only the PCRs of Bulgaria and Austria refrain from storing negative information. Overall, the table highlights the disperse structure of credit reporting mechanisms across Europe. Some authors interpret this as a need (and also a challenge) for further harmonisation, both to react to higher mobility across national borders and in an endeavour to create a single European credit market (see also Jentzsch 2007).

Some more stylised facts not visible from the table: With regard to the coverage of credit reports, the UK, Germany and Sweden have the highest number of credit reports per person (Jappelli & Pagano 2002). In general it is known that the number of credit reports per capita is positively associated with household mobility. This is plausible as the benefit of shared information is greatest when lenders borrow to large numbers of unknown customers.

	Public credit registers			Credit bureaus		
		information collected			information collected	
Country	established in	positive	negative	established in	positive	negative
Belgium	1967	Yes	Yes			
Bulgaria	1998	Yes	No	1995	Yes	Yes
Czech Republic	1994	Yes	Yes	2000	Yes	Yes
Denmark				1971	No	Yes
Germany	1934	Yes	Yes	1927	Yes	Yes
Estonia				2001	Yes	Yes
Ireland				1963	Yes	Yes
Greece				1993	No	Yes
Spain	1962	Yes	Yes	1967	No	Yes
France	1946	No	Yes			
Italy	1962	Yes	Yes	1989	Yes	Yes
Cyprus				2001	Yes	Yes
Latvia	2003	No	Yes	unknown	Yes	Yes
Lithuania	1996	Yes	Yes	2000	No	Yes
Luxembourg						
Hungary				1990	Yes	Yes
Malta				2002	No	Yes
Netherlands				1965	Yes	Yes
Austria	1986	Yes	No	1941	Yes	Yes
Poland				2001	Yes	Yes
Portugal	1978	Yes	Yes	1996	Yes	Yes
Romania	1999	Yes	Yes	2000	Yes	Yes
Slovenia	1994	Yes	Yes	unknown		
Slovakia	1997			2003	Yes	Yes
Finland				1961	No	Yes
Sweden				1890	Yes	Yes
United Kingdom				1960	Yes	Yes

Table 3.2: Credit reporting systems in the EU member states as of 2005/2006[26]

3.2.2 Credit Scoring as an Instrument of Credit Reporting

The "Guide to Credit Scoring" issued by a number of banks in U.K. describes credit scoring and its purposes as follows[27]:

"Credit scoring measures the statistical probability that [a] credit will be satisfactorily repaid. It is based on the fact that it is possible, using statistical techniques, to predict the future performance of applicants with similar characteristics to previous applications (either of the credit grantor itself or groups of credit grantors).

[26] Source: Jentzsch (2007) based on World Bank (2003) and Jappelli & Pagano (2002).
[27] See http://www.bba.org.uk/content/1/c4/66/11/Guide_to_Credit_Scoring_2000.pdf, p. 4.

Scoring calculates the level of risk and reduces the element of subjectivity in lending decisions. It enables credit grantors to manage their business more effectively. This benefits the majority of customers who wish to borrow only what they can afford to repay."

Credit scoring has already been described and discussed in many FIDIS deliverables, especially with a legal focus in (Hildebrandt, Backhouse 2005: 73-74), (Hildebrandt, Gutwirth 2007: 203-211) and (Hildebrandt, Kindt 2009). In addition, a study in Germany commissioned by the German Federal Ministry of Consumer Protection, Nutrition and Agriculture was published by Kamp and Weichert (2005).

Technically, scoring can be categorised as non-distributive group profiling of the type "classification". An object of investigation (in this case a customer asking for a credit), described by various attributes, is assigned to a certain pre-specified class (in this case a certain defined class of credit risks).

One example of credit scoring operated in Germany is a service of the so-called "Schutzgemeinschaft zur Sichererung des Kreditwesens (Schufa)". In this case categories used are scoring values ranging from 0 to 1000, representing the likeliness of credit failure. These score values are mapped to 12 rating levels which are used in the decision process for granting a credit (Kamp, Weichert 2005: 33). Also attributes can be concluded (Kamp, Weichert 2005: 50 – 56) and the scoring algorithm used in this case is documented (Kamp, Weichert 2005: 58; in this case a logistic regression algorithm is used), though the parameterisation of the algorithm is considered to be a trade secret by the Schufa. The generic process of credit scoring is shown in the following figure:

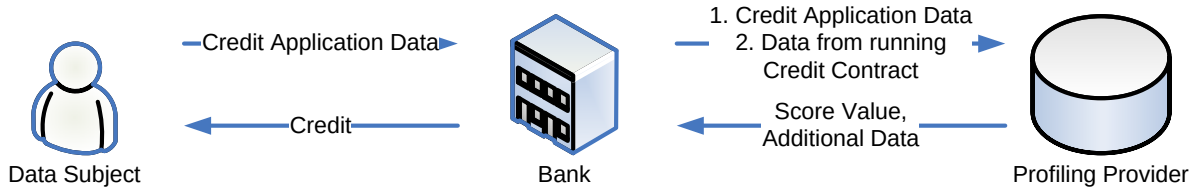

Figure 3: Generic process of credit scoring

In the context of the modernisation of the Federal Data Protection Act currently a discussion on the balancing of trade secrets and transparency requirements based on data protection is ongoing in Germany (see also Hildebrandt, Kindt 2009). An updated report on this discussion can be found in chapter 5.3 in this deliverable.

3.2.3 Empirical Evidence on the Effect of Credit Reporting

Jappelli and Pagano (2002) find that bank lending is higher and credit risk is lower in countries where lenders share information, regardless of the private or public nature of the information sharing institutions. They conclude this from a cross-country study that measures bank lending as ratio between total bank claims to the private sector and GDP (as of 1994/1995). Credit risk is measured with a survey-based indicator collected for the

International Country Risk Guide survey of leading international bankers. More precisely, the results suggest that information sharing is associated on average with an increase of bank lending per GDP by about 20 percentage points. (A similar quantitative interpretation of the effect on credit risk is not meaningful, due to the artificial scale of the survey responses.) These relations even exist when taking into account a number of possibly alternative explanations, including GDP growth and size as well as indicators for rule of law and creditor rights. Jappelli and Pagano (2002) further report that the effects are stronger in countries where positive information is shared in addition to basic negative information. However, they could not find any statistically significant difference between the effect of PCRs and private credit bureaus and conclude that both can be regarded as mutual substitutes. These empirical facts are independently confirmed in similar investigations by the World Bank (2003) and, as a side-result, in an empirical analysis of the effect of privacy regulation on credit reporting by Jentzsch (2007, Appendix Table 5.16, p. 289).

3.3 Prevention of Systemic Risks

Kaufman and Scott (2003) define *systemic risk* as the risk of a breakdown in an entire system, as opposed to breakdowns in individual parts or components. Systemic risk may occur in various parts of the financial sector. In banking, it refers to the clustering of bank failures in a single country, or, as recently witnessed, throughout a globalized financial system. In securities markets,[28] systemic risk is caused by high correlation of asset prices and materializes in simultaneous declines of a large range of securities.

By the very nature of systemic risk, it is impossible to assess it solely by the analysis of its individual components. For example, it is not sufficient to measure the quality of individual loans in a banking book, since systemic risk emerges from the interrelation between loans. If loans are sufficiently diverse, i.e., if their probability of default is independent of each other, then individual losses are balanced in the portfolio and systemic risk is low. Contrary, if all loans depend on a single factor (e.g., economic success of a large employer in the region) and if this factor turns adverse, all loans default *at the same time*. In this case the lender has to absorb high losses due to systemic risk. The same can be said for (the dependency on) the level of individual lenders that constitute the financial sector in an economy.

Systemic risk in banking needs special attention from regulators due to the distinct function of banks as intermediaries in a market economy. Therefore policy makers are keen to avoid bank failures, which typically go along with high economic cost in both fiscal terms and forgone potential production (e.g., Hoggarth and Saporta 2001). It is argued that credit registers are the right tool for banking supervisors to monitor systemic risk as they reveal the structure of and links between many lenders as well as the distribution of risk (Artigas 2004). However, complete coverage is crucial. This is why public credit registers (PCR) combined with legal provisions of mandatory reporting are seen as the solution of choice with regard to the ability to monitor systemic risk. Only a few PCRs in Europe were established with the aim to prevent systemic risks. These include Germany's *Millionenkredit-Evidenzzentrale*, which contains

[28] A security is a fungible, negotiable instrument representing financial value. Securities are broadly categorized into debt securities (such as banknotes, bonds and debentures) equity securities, (such as common stocks) and derivative (finance) contracts (such as forwards, futures, options and swaps).

data on all loans granted by German financial institutions above a relatively high threshold of 1.5 million euro (Jentzsch 2007), and Spain's *Central de Información de Riesgos* with a lower threshold of 6,000 euro (Artigas 2004).

Banking supervisors can use credit registers to support both on-site inspections and to carry out off-site monitoring of credit and concentration of risk. In both cases, the sheer amount of information in the register calls for profiling techniques to detect patterns and monitor their evolvement over time. Unfortunately, there exists little literature[29] on the profiling techniques actually employed, nor on measurable benefits of credit register data to support supervisory activities. Further uses of the PCRs have been conceived in the context of Basel II, where credit information could be useful to validate the calibration of bank-internal rating models (Artigas 2004). However, this is generally a tedious task for supervisors and as such becomes increasingly criticised in the light of the recent financial crisis (Gerding 2008).

In the aftermath of the collapse of *Long Term Capital Management* (LTCM), a hedge fund, PCRs have been proposed as tools to monitor systemic risk created by the exposures of highly leveraged institutions (HLI), including hedge funds (Basel Committee on Banking Supervision 1999). However, no agreement on the required international level could be reached. The same idea has been taken up in the 2009 Issing report,[30] which coined the term "global risk map". The report was commissioned by the German Federal Government to prevent future financial crises.

3.4 Anti-Money-Laundering / Prevention of Terror-Financing

Anti-money-laundering[31] is regulated on a European level based on the Directive 2005/50/EC on the prevention of the use of the financial system for the purpose of money laundering and terrorist financing.[32] Financial and also some non-financial institutions are obliged to ongoingly monitor financial transactions and to identify potential customers (individuals as well as organisations) carrying out money laundering.[33] For this purpose profiling is applied, though methods and algorithms used are not publicly available. From the information available it can be concluded that money transactions and socio-demographic data from customers and account holder is used in the process.

In case of suspicious activities a so-called Suspicious Activity Report (SAR) needs to be drafted and submitted to the national responsible authority, called the Financial Intelligence Unit (FIU).[34] The FIU coordinates further criminal investigations and international

[29] Available literature is mostly included in this report, other knowledge comes from working experience of the authors.
[30] http://www.bundesregierung.de/nn_1272/Content/DE/Artikel/2009/02/2009-02-09-bk-bmf-issing-bericht.html
[31] Anti-money-laundering already was presented and discussed in a previous FIDIS deliverable (Hildebrandt, Backhouse 2005: 57).
[32] See http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2005:309:0015:0036:EN:PDF
[33] cf. Art. 8 of Directive 2005/50/EC.
[34] For data protection issues in regard to the activities of FIUs see the Report of the EU FIU Platform. Download: http://ec.europa.eu/internal_market/company/docs/financial-crime/fiu-report-confidentiality_en.pdf

information exchange. The generic process of anti-money-laundering is shown in the following figure:

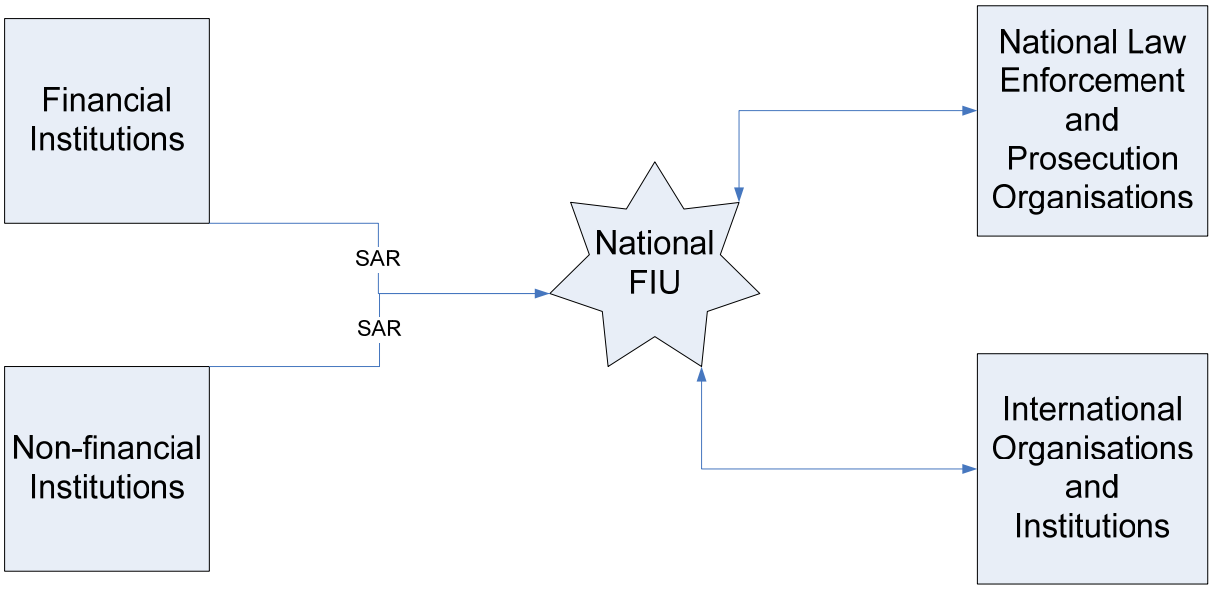

Figure 4: Generic Process of anti-money-laundry[35]

As the assumptions used for the profiling are simplified and rely on socio-demographical group profiles, certain types of cases of money laundering are not easily uncovered (see Hildebrandt, Backhouse 2005: 58). The next chapter will deal with the implementation of anti-money-laundering and anti-fraud programs in more detail. The focus will be on the practical problems of implementing these strategies.

[35] Figure based on Canhoto, in (Hildebrandt, Backhouse 2005: 58)

4 Implementation and Practical Problems

As described in the previous chapter, financial institutions are implementing profiling strategies in different areas with different purposes. In this chapter we will discuss the practical implementation of such profiling strategies, indicate where practical problems arise and, finally, indicate which advantages and disadvantages may arise for the parties involved.

We will not do this for all cases described in the previous chapter, but, instead, we will focus on one particular case: KYC. KYC stands for Know Your Customer and is a profiling process that is mandatory for most banks in Europe and the United States. It is aimed at finding fraud, terrorism funding and money laundering. Due to new legislation (particularly in the United States), financial institutions, particularly banks, have to find out with whom they are doing business. Of each client a risk profile has to be made and in cases of high or unacceptable risks action can be taken, for instance, by removing clients from the lists of business partners or by informing the supervisory authorities.

4.1 Implementation

In practice, implementing a KYC policy means for a financial institution that there has to be an organized a system that builds risk analyses of all existing and new clients. In order to determine the scope of a KYC project, it is necessary to know how many clients there are. Since multinationals may have many clients (up to hundreds of thousands or millions of clients) this may cause difficulties. Once a client has been identified, the risk analysis can be started. Usually a risk analysis consists of mapping a number of characteristics of a client, collecting the evidence for these characteristics, and, finally, attaching a risk index to this by weighing the characteristics.

Client characteristics that can be thought of in case of natural persons are name, address, date of birth, solvency, number and types of accounts, data on fraud or criminal activities in the past, etcetera. Evidence regarding the identity usually consists of copies of passports. Evidence for other characteristics may be government documentation, such as statements of good behaviour.

Characteristics of legal persons may consist of name, address, date of incorporation, business activities, names of directors, names of shareholders, names of owners. Furthermore, it may be investigated whether a particular legal person is registered at or supervised by a stock exchange, a local chamber of commerce, a financial authority, a local government, or, any other supervisory authority.

Proving the identity of a legal person may be done with a certificate of incorporation or a registration at the chamber of commerce. Other characteristics may be proven with the use of documentation of chambers of commerce and supervisory authorities, annual reports with accountancy statements, certificates of incorporation, copies of passports of directors, shareholders and owners, etcetera.

Determining the risk is a final weighing of all the characteristics of a particular client. Several characteristics may indicate increased risks:

- *The location of the client:* countries such as Iraq, Somalia or Libya are considered high risk because there is little or no supervision on natural persons and legal persons. The same, but to a lesser extent, applies to countries like Russia and India. Furthermore, the US government prohibits trade with particular countries. Examples are Cuba and Iran (US Sanctions list, 2006).
- *Business activities:* particular business activities are sensitive to money laundering and terrorism funding. Examples are casinos, exchange offices, and diamond trading offices.
- *Legal company structure:* so-called shell companies are administrative constructions, where no real business activities are performed. Due to favourable tax climates, these constructions are often not very transparent when determining who the directors or the owners are. Many shell companies are found in favourable tax climates, such as on the Cayman Islands, the British Virgin Islands, or, Bermuda. Other legal constructions may also lack transparency. Examples, though depending on the legal regime of a particular country, may be foundations and structures with silent partners.
- *Presence at black lists:* when directors, shareholders or owners appear on black lists, this may indicate increased risks. In case of legal persons, there are also black lists with company names. It is important to distinguish lists with increased risk and lists that prohibit transactions with particular clients. Sometimes lists with increased risk are indicated as 'grey lists' to distinguish them from the 'real' black lists that contain prohibitions.

The latter issue, black lists, may need some further elaboration. Both in the United States and in the European Union several black lists are used. An example is the OFAC list (OFAC List, 2006) of the Office of Foreign Assets Control (OFAC) of the US Department of the Treasury. On the OFAC list are more than 5000 persons the US government has marked as terrorists and/or heavy criminals. Doing any business with persons on this list is prohibited. Examples of other lists that are being used to check persons are the FBI ('most wanted') list, the EU-list of terrorist organizations, the Australian DFAT list, the Bank of England list, lists of Europol, and many other lists. Apart from lists of suspects of terrorism and criminality, it is also possible to check against other lists, such as solvency lists.

Apart from characteristics that increase the risk, there are also characteristics that decrease the risk. This is often the case when independent authorities are supervising a client. For legal persons, a listing on a stock exchange may only be possible when higher demands regarding the transparency and a solid financial situation are met. In several countries registration at the chamber of commerce is mandatory and subject to critical investigation. Clients operating in financial markets, such as banks and insurance companies, are usually subjected to critical supervision. Clients that are part of or related to (semi)government organizations are also supervised in many cases. Obviously these characteristics only decrease the risk in countries where the government and supervisory authorities are considered reliable.

Based on a weighing of all characteristics that have been discovered in the process, a risk assessment is made. In cases of increased risks, this may involve periodical scrutiny. In cases of unacceptable risks it may be decided to end relationships with such a client.[36] Obviously a

[36] This is not unrealistic. For instance, American Express cut customer credit lines based on profiles (not on blacklists), see Lieber (2009).

risk profile, once it has been created, has a limited durability and will have to be assessed on correctness periodically. Both the data in the profile and the weighing and risk assessment may then need to be updated.

4.2 Practical Problems

In practice, implementing the legislation into a process as described above may lead to several problems.

Unclear Scope
Large multinationals do not always know how many customers they have. This is often due to their growth by mergers and acquisition. When client databases are kept in different information systems, it may be difficult to couple these databases. There may be dozens of databases that may contain hundreds of thousands or millions of clients. As a result, there may be a fragmented registration of clients and a lot of overlap in the data. For instance, a person or company may be a client at several banks that were merged. Due to overlap it may be hard to determine how many unique clients are hidden in the various information systems. When it is unknown how many customers are to be analyzed, and which clients these are, the scope of the project is unclear.

Identification Issues
Identifying persons or companies may be difficult. How do you know whom you are dealing with? When a particular database mentions Mr. William White and another database mentions Mr. Bill White, it may be the same person or a different person. When the address is the same in both records, it is likely that it is the same person using a shortened version of his name. The probability that it is in fact the same person increases when more characteristics are identical, such as data of birth, phone number, social security number, etc. Currently there are technological solutions that may establish, based on overlap, whether it is the same identity that is being dealt with in such cases. An example is Entity Analytics Solutions of IBM.

When dealing with companies, this may be even more difficult, since legal structures of large companies often contain many different legal persons. For instance, Custers PLC, Custers International, and Custers International Holding PLC may all be different companies and different legal persons. However, it is also possible that these names refer to the same company, with the same directors. A company may use different names for branding and marketing purposes. It may be that the official name registered at the chamber of commerce is much longer than the name used for advertising. Often companies use different names for brands or products. Names of divisions may also differ from the conglomerate name.

Persons behind Organisations
Obviously identifying companies is not a goal in itself. The goal behind this is to find out the persons behind organizations, such as directors and shareholders. Sometimes the shareholders of companies are other companies. Searching for a parent company may lead to natural persons who are directors and shareholders. Of all persons involved in a company it should be investigated whether they are related to terrorism or money laundering.

However, it may be difficult to find the persons behind organizations. Many international companies have parent companies in countries other than the country where a subsidiary is located. As a result, the search may depend on other sources (such as local supervisory authorities and chambers of commerce). These other sources may be in different languages and subject to other rules. Many companies are located in countries with strict banking secrecy for tax purposes (e.g., Luxemburg, Switzerland) or in countries with favourable tax climates that are not very transparent (e.g., Cayman Islands, British Virgin Islands, Channel Islands). Other company structures, such as foundations or partnerships with silent partners may also lack transparency; this may vary from country to country.

All names that are found must be checked against the black lists that are used by secret services and surveillance authorities such as the CIA, FBI, Interpol, and, Europol. More general checks, such as bad press, may provide more background information. Note that all these checks against black lists may only result in 'hits' on persons that were in the past related to terrorist incidents. First time terrorists intending to prepare, finance or commit an attack, are usually not on black lists.

Standardisation
KYC legislation states that all clients must be profiled. For large international financial institutions, this may involve hundreds of thousands of clients. Because of the amounts of time and money involved, there is a tendency towards standardization. Procedures are required to streamline the processing of large amounts of data. However, standardization and procedures usually focus on the average funds, whereas tracing terrorist funds should focus on the exceptions. Using a standardized and predictable approach may have as a result that the suspects may be overlooked. Furthermore, there is the risk that the persons that do not want to be traced have plenty of time to change their strategies in order to avoid being labelled with increased risk profiles.

Documents rather than Persons
Although a KYC policy aims at finding suspect persons, the current profiling strategies are performed on the basis of documents. The main reason for this is usually not to bother clients with requests for information. This is, however, an indirect type of checking, because the integrity of the document is checked, rather than the integrity of the person. This means that two things can go wrong: the integrity of the document can be messed with, or the link between document and person can be messed with.

The first problem, messing with the documents, occurs regularly in international criminality and terrorism. Persons may use different passports and aliases. For this reason, documents are nowadays equipped with characteristics that are hard to forge, such as graphics, watermarks, holograms and seals. Distinguishing real and fake documents requires frequent training of inspectors on these increasingly complex characteristics.

The second problem, messing with the link between person and document, occurs more and more frequently. For this reason, passports of many countries are nowadays equipped with biometrics. Obviously this is not possible for other documents, such as documents regarding legal persons. By integrating body characteristics of a person in the identity document, the link between person and document can be strengthened. This makes it more difficult for people to hide behind documents. Note that these do not have to be fake documents. A person

may simply use a real document of another person; a setup known as look-alike fraud. The link between person and passport may be hard to verify. Inspectors often focus on the picture in the document, but the passport photo may be old. A beard may have been shaven or glasses may have been replaced with lenses.

4.3 Pros and Cons of Profiling

Group profiles are usually produced and used with particular purposes in mind. The user of group profiles expects certain advantages when these goals are reached. For financial institutions, these purposes are usually risk management, credit management, fraud control, anti-money laundering, etc. However, these advantages may be disadvantageous for others. Although most of the advantages and disadvantages are dependent on the context and the perspectives of persons (e.g. the profiler/profiling institutions and the person/client/organisation being profiled), there are some general advantages and disadvantages of profiling, which are dealt with in this section.

4.3.1 Advantages

The advantages of group profiling usually depend on the context in which they are used. nevertheless, some advantages may hold for many or most contexts. Some of the advantages of group profiling are pointed out in comparison with individual profiling or no profiling at all. The main advantages concern *efficacy*, i.e., how much of the goal may be achieved, and *efficiency*, i.e., how easily the goal may be achieved. Group profiling may process huge amounts of data in a short time; data that is often too complex or too much for human beings to process manually. When many examples are present in databases, (human) prejudices as a result of certain expectations may be avoided.

Group profiling may be a useful method of finding or identifying target groups. In many cases, group profiling may be preferable to individual profiling because it is more cost efficient than considering each individual profile. This *cost efficiency* may concern lower costs in the gathering of information, since less information may be needed for group profiles than for individual profiles. But higher costs may also be expected in the time-consuming task of approaching individuals. While individuals may be approached by letter or by phone, groups may be approached by an advertisement or a news item.[37]

Another advantage of group profiling over individual profiling is that group profiles may offer more possibilities for selecting targets. An individual may not appear to be a target on the basis of a personal profile, but may still be one. Group profiles may help in tracking down potential targets in such cases. Such selection may also turn out to be an advantage for the targets themselves. For instance, members of a low-risk group may be applicable for special offers and reductions. But selection may also be unwanted or unjustified, in which case it may be a disadvantage for the target.

[37] Obviously, these advantages do not only apply to the financial sector, but to other sectors (e.g., advertisement) as well.

Group profiling may be more useful than no profiling at all. Without any profiling, without any selection, the efficiency or "hit ratio" is usually poor. For instance, advertising using inadequately defined target groups, like on television, is less efficient than advertising only to interested and potentially interested customers.

4.3.2 Disadvantages

The use of group profiles may have negative effects as well. Note that the disadvantages and risks mentioned below are closely connected with the advantages mentioned in the previous subsection. Sometimes, an advantage from one perspective may involve a disadvantage from another perspective. The effects described below may lead to unwanted intrusions and/or perceived loss of privacy and autonomy for the person that is profiled..

One of the most obvious risks of group profiles is that they may be used as selection criteria in a way that is considered unjustified by group members or others. This risk is particularly present when non-distributive group profiles are used. *Selection* is one of the main applications of group profiles. Selection of fraudsters may be beneficial for most stakeholders, but selection may also be used for more controversial issues, such as refusing loans and mortgages.

Some of the group profiles constructed by financial institutions or others may become 'public knowledge', which may lead to the *stigmatisation* of that particular group. There are many examples of stigmatisation and these examples are not always limited to the ones usually mentioned in human rights documents, such as sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, or birth.[38]

Another issue is *confrontation*. When group profiles are used for preventive purposes, it is possible that people in risk groups be approached with a warning. In this way, people are confronted with their prospects, without having requested to be given such information about themselves. For instance, if there is a correlation between creditworthiness and life expectancy, people may indirectly be informed about the latter. Especially in the case of very negative group profiles, such a confrontation may have a large impact on people's lives. In some cases, people may prefer not to know their prospects.

Another type of risk involves one-sided information supply. This may be caused by *customisation* or *personalisation*,[39] through which companies try to approach people (customers) in a manner that corresponds with their personal preferences. According to most companies, good service typically includes giving a great deal of attention to a customer and

[38] These are the criteria explicitly mentioned in Article 14 of the Convention for the Protection of Human Rights and Fundamental Freedoms, prohibiting discrimination.

[39] Customisation and personalisation are sometimes used as different concepts, see Treiblmaier H. et al. (2002). *Customization* requires users to explicitly control the adaptation process' whereas *personalization* is driven by the secondary (mis)-use of traces that people leave behind while surfing on the Internet. As such, personalization can occur without the consent and even without the awareness of citizens. Some argue that the term 'impersonal personalization' suits the practice better (e.g. Won, 2002:31). In this meaning, de-individualisation is particularly related to personalisation.

trying to fulfil his needs and wishes. Although customisation may lead to a perceived better service, it may also lead to one-sided information supply.

Although it may seem that group profiles lead to a more individual approach (e.g., by customisation), the use of group profiles may, in fact, lead to *de-individualisation*. This is a paradox. Group profiles result in a tendency to judge and treat people on the basis of their group characteristics instead of on the basics of their own individual characteristics and merits.[40] Thus, the use of profiles is likely to lead to a more one-sided treatment of individuals. Besides, individuals may be given an identity that is not of their choosing.[41] Note that this may also be the case for personal profiles. Still, the negative effects in the case of group profiles may be larger because of non-distributivity, when the characteristics ascribed to group members may not be valid for them as individuals.

[40] Vedder (1999), Bygrave (2002), p. 312.
[41] Bygrave (2002), p. 291.

5 Legal Regimes

In this chapter, the legal regimes applicable for profiling (in relation to financial issues) are described. First, privacy and data protection are discussed, second, intellectual property rights are discussed. This chapter concludes with a note on the current debate on credit scoring in Germany and a note on the limitations of privacy preserving data mining.

5.1 Privacy and Data Protection

Although most issues regarding profiling are closely related to data protection, they are often not privacy problems. Privacy contains many different aspects, including spatial privacy, relational privacy, communicational privacy and informational privacy. Personal data protection focuses on informational privacy. This is where most legislation is available.

The main legal regimes of interest concerning data protection law in the European Union are the Treaty of Strasbourg,[42] the European Data Protection Directive[43] and the national Personal Data Protection Acts. The central rules of these data protection laws are based upon and embody a set of principles that was drafted by the Organisation for Economic Co-operation and Development (OECD) in 1980, the so-called *privacy principles*.[44] The principles are explained below. In all these laws, these central rules apply to a particular set of data, namely, *personal data*, i.e., any information relating to an identified or identifiable natural person. In this section, it will be discussed how this concept of personal data may raise problems regarding data protection.

5.1.1 Privacy Principles

In 1980, a set of principles for fair information processing was developed by the Organisation for Economic Co-operation and Development (OECD). The OECD is an organisation of 30 countries world-wide that is committed to democratic government and a market economy that works on economic and social issues.[45] The principles developed by the OECD, commonly referred to as *the privacy principles*, are [46]

- the *collection limitation principle*, stating that "[t]here should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject";[47]
- the *data quality principle*, stating that "[p]ersonal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete and kept up-to-date";

[42] Council of Europe, Convention no. 108, January 28th 1981.
See http://www.coe.fr/dataprotection/Treaties/Convention%20108%20E.htm
[43] European directive 95/46/EG of the European Parliament and the Council of 24th October 1995, [1995] OJ L281/31. See also http://europa.eu.int/eur-lex/en/lif/dat/1995/en_395L0046.html.
[44] See also Bygrave (2002), p. 2.
[45] See http://www.oecd.org.
[46] http://www.oecd.org/document/18/0,3343,en_2649_34255_1815186_1_1_1_1,00.html
[47] This principle is sometimes referred to as the *principle of minimality*, see Bygrave (2002), p. 341.

- the *purpose specification principle*, stating that "[t]he purposes for which personal data are collected should be specified … and that the data may only be used for these purposes";
- the *use limitation principle*, stating that "[p]ersonal data should not be disclosed, made available or otherwise used for purposes other than those specified, … except a) with the consent of the data subject; or b) by the authority of law";
- the *security safeguards principle*, stating that reasonable precautions should be taken against risks of loss, unauthorised access, destruction, et cetera, of personal data;
- the *openness or transparency principle*, stating that the subject should be able to know about the existence and nature of personal data, its purpose, and the identity of the data controller;
- the *individual participation principle*, stating, among other things, that the data subject should have the right to have his personal data erased, rectified, completed, or amended;[48]
- the *accountability principle*, stating that the data controller should be accountable for complying with measures supporting the above principles.

The former four principles focus on the data and the conditions under which processing of the data is allowed, and the latter four principles are duties of those responsible for the processing of personal data and rights of the data subjects.

5.1.2 Personal Data

Personal data is defined as any information relating to an identified or identifiable natural person.[49] This definition may be found in the OECD guidelines mentioned before, but also in the Treaty of Strasbourg, the European Data Protection Directive and the national Personal Data Protection Acts. Note that this definition explicitly excludes data concerning legal persons, non-identifiable persons, and deceased persons. In the case of group profiling, this definition is important for determining whether the information in group profiles is personal data or not, and, consequently, whether or not data protection laws are applicable.

A difficulty with the concept of personal data is that it may not be clear which data can be used to identify a natural person and which data are not.[50] Identifiability is often considered an absolute standard: data are either identifiable or not. It is sometimes suggested that this means that only in cases where *every possibility* of linking data to an identified person has vanished is the data not considered to be personal data. However, with technology providing more and more means of linking data, and with increased dissemination of data, such an interpretation may result in almost all data being considered personal data.

Because it is impossible to check if every possibility of linking data to an individual has vanished, an element of reasonableness is often included in the concept of identifiability: when it would require a disproportionate amount of time, money, and manpower to identify

[48] Note that, in the European Data Protection Directive this principle applies only to incomplete or inaccurate data, or data that are irrelevant or processed illegitimately.
[49] This person is referred to as the *data subject*.
[50] Especially with combinations of data, this may be difficult.

the data, these data are not personal data.[51] However, with the possibilities of information and communication technologies increasing further, the amounts of time, money, and manpower needed to identify data become steadily smaller. Methods of coupling databases may be especially efficient in identifying data. Thus, even with such an account of reasonableness, it seems that more data will be comprised in the concept of personal data as time passes.

5.1.3 Applicability

Bygrave argues that the definition of personal data as "any information relating to an identified or identifiable natural person" may be read as two cumulative conditions; namely, that the data must facilitate the identification of such a person and that the data must relate to or concern a person.[52] If either the condition of identifiability or the condition of a data-person relation is not fulfilled, the data are not personal data, and the personal data protection is not applicable.

Let us start with the condition of identifiability. As already discussed in the previous subsection, an element of reasonableness was included in the concept of identifiability: only when it would require a proportional amount of time, money, and manpower to identify the data may these data be considered personal data. However, with the increasing possibilities of information and communication technologies, the element of reasonableness may become increasingly easier to fulfil. It may be argued that, when more data are considered personal data, this may extend the applicability of the privacy principles in data protection law. For the data subject, this may be an advantage because the privacy principles tend to protect the position of data subjects. For the data controllers, this may be disadvantageous, because extensions in the applicability of data protection laws may hinder them in collecting and processing data. In order to provide solutions for the possible problems of data mining and group profiling, it may be argued that it is not necessary for data protection law to impose conditions on the use of increasing amounts of data, but rather to impose conditions on the use of data that may hinder meeting the needs of those involved.

Another problem with the concept of identifiability is that it suggests that data are either identifiable or not. However, from a practical/technological point of view, there are different degrees of identifiability. Non-identifiable data may still be used for group profiling, as long as the data are not absolutely anonymous. Thus, the concept of identifiability does not reflect the practical situation. From a technological perspective, preventing the use of particular forms of profiling requires focusing on linkability rather than on identifiability (Custers, 2004).

When the condition of identifiability is applied to group profiles, it may be argued that group profiles generally do not contain data about identified individuals: deriving individual data from group data requires an inference step to be made first. The argument that group profiles may become personal data when inferred to individuals may be true, but such inference does not always take place.

---

[51] Note that for identifiability it is necessary to determine whether or not identification is *possible*, not whether or not identification will in fact take place.
[52] Bygrave (2002), p. 42.

The second condition for personal data, a data-person relation, may also be difficult in group profiles. Group profiling may ascribe to people characteristics that lack integrity.[53] Group characteristics may not be valid for individual members of a group (non-distributivity). Thus, it is to be doubted if such incorrectly ascribed data should be considered personal data. Since personal data are "any information *relating to* an identified or identifiable natural person", the question is whether data are ascribed in such a way that they *relate to* the particular individual. For instance, if a group profile based on a zip code ascribes to 25-year-old female students the properties "retired" and "male", should this be considered personal data with many errors or should it simply not be considered personal data? It may be argued that once characteristics are ascribed to identifiable individuals, these characteristics, whether correct or not, relate to them. However, when characteristics are ascribed to groups of people, it may be more difficult to maintain this argument and the personal data protection may not be applicable.

The concept of personal data limits the applicability of data protection law with regard to group profiling. Only in the first and the last step of the KDD process (see Figure 2.2) may personal data occur.

5.1.4 Protection

When personal data protection is applicable, the question may be raised, how much protection it offers. The protection offered consists mainly of the privacy principles that form the central rules of personal data protection. These principles of fair information processing focus on the position of the data subject. The privacy principles do not prevent the use of group profiling, but impose conditions on the ways in which information is processed.

A data subject may consider treatment on the basis of group profiles to be unfair or incorrect (or both). When a data subject considers himself unfairly treated on the basis of a group profile, personal data protection acts may offer possibilities of liability and redress. Incorrect treatment, however, results from the use of unreliable data, and it may be suggested that, using the individual participation principle the data subject can use his right to have incomplete or inaccurate data changed.[54] Although this right may not be denied, it may not offer protection against incorrect treatment. This problem is illustrated in Figure 5.1 and is discussed below. Starting with wrong data, but giving a person the right to amend the data, resembles a guilty-until-proven-innocent system.[55] It may be argued that such infringements of the presumption of innocence should be avoided under the rule of law.

According to the individual participation principle, a person has the right to have incomplete or inaccurate data changed, destroyed, or removed. The right to have data changed or

[53] Errors are common. For instance, a study by the U.S. Public Interest Research Group of credit-report accuracy and privacy issues found that 29 per cent of credit reports contain serious errors that could result in the denial of credit, loans, or jobs, and that altogether 70 per cent of credit reports contain mistakes. Public Interest Research Group (1998).

[54] Note that the data controller also has the obligation to keep the data accurate, complete, and up to date on the basis of the data quality principle.

[55] See also Bing (1986), who states that by being a member of a group the burden of proof may be reversed.

removed may appear clear on paper, but may prove to be difficult to assert in practice.[56] If the data collector ignores the openness principle, it may be difficult to find out what data have been collected and processed. If the openness principle is applied, the data subject may ask to have his personal data removed, but the data collector may ignore this request (enforcement is discussed below) or remove the identifiers of the personal data so that the personal data protection is no longer applicable. Finally, if the data collector is willing to change or remove the data, this may prove difficult to do. As errors spread throughout the system and accumulate as data are disseminated and merged, data subjects may find themselves affected by the same errors over and over again. This problem may become even greater when other organisations are involved. A data collector who has sold several copies of his data to other companies may inform these companies about the changes, but has no obligation to do so. Alternatively, the data subject may contact each of these companies with his request to change or remove his personal data, which means starting all over again; see Figure 5.1.

Another problem may arise when the security system of the database involves different access levels. A person who has access at a low level is not authorised to change data at all levels. Thus, only in a security system where a person has access to all security levels would it be possible to perform adequate changes in a database. This problem is called *poly-instantiation*.[57]

The rectification of incorrect data may involve the release of more personal data, which may result in an experienced loss of (informational) privacy.[58] When characteristics are ascribed incorrectly, a request to change the incorrect data may imply that the data subject needs to provide the correct data, in order not to be judged on the incorrect data. Such a release of personal information may imply that the data subject experiences a loss of (informational) privacy. Custers refers to this as *the privacy paradox*.[59]

[56] It may be argued that this is a problem of enforcement, but when difficulties of enforcement are to some extent inherent in legislation, they may be regarded as a shortcoming in that legislation.
[57] Denning, D.E. (1988) Lessons learned from modelling a secure multilevel relational database system. In: *Database Security: status and prospects*, C.E. Landwehr (ed.) Elsevier science publishers, IFIP.
[58] Note that such an experienced loss of privacy may occur without any infringement of a legal right.
[59] Custers (2004). Note that others have already used the term privacy paradox for other situations. See, for instance, Etzioni (1999), who used it to indicate that, although the government (Big Brother) was traditionally seen as the biggest threat to privacy, it may nowadays be to the same government that we need to turn for protection from the threats that private companies (Big Bucks) pose to our privacy.

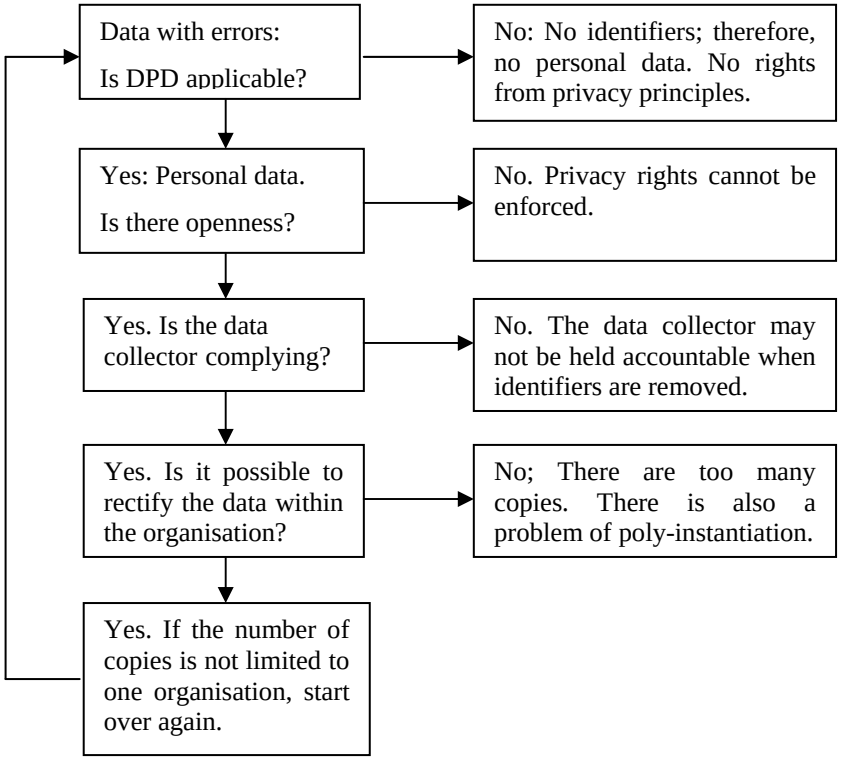

Figure 5.1: *Rectification problems and other problems of applying the individual participation principle.*

It could be argued that destruction of data might be preferred to rectification of data in such cases. However, in the case of positive characteristics, destruction of the data may, for instance, result in a data subject not getting particular special offers he would have liked to receive. In such cases, a data subject has to determine the risks and benefits involved in releasing his data. In extreme cases, group profiling may become a way of collecting personal data because of the privacy paradox: companies may start *rough profiling* based on limited or inaccurate data. They may explicitly use the rough profiles combined with some invitation to rectify the incorrect data in order to obtain the correct data.

It may be difficult to hold a data collector responsible for not complying with the openness principle when it is not known who is responsible for what. Making it attractive for a data collector to pursue openness may help in getting out of this impasse. This may be done by making a general policy of openness a competitive argument.[60]

According to general economic theory, markets function efficiently only when well-informed consumers are able to choose from among competing products or services. This means that the above-mentioned competition would only work if there were a real choice for the customer. If all financial institutions required the filling out of extensive forms with personal information, there would not be such a choice. A second requirement is that there is some awareness among the customers about the privacy issues involved. Self-regulation would

[60] A policy of openness is usually part of a *privacy policy*. Note that openness may also be harmful for competition positions as the revelation of strategies and openness adversely affect competition positions when data subjects do not like particular forms of data processing.

mean, in theory, that if privacy is important to consumers, then organisations will respond to the perceived consumer demand and will provide privacy protection. But many people are not aware of the possible violations of their privacy and thus do not think about the consequences of giving away their personal data, for instance, when applying for an e-mail address.[61] When asked, people may say they care about privacy, but they might not always realise how the linking of data may result in a gradual erosion of privacy. Profiling techniques, by their nature, tend not to be visible processes for data subjects.[62]

As a final remark, it should be mentioned that accountability might be difficult to enforce in the international collection and processing of data. Data protection law may not deter criminals who commit crimes using computers, as it is possible with today's networks to operate over large distances and users can remain untraceable, making it difficult to enforce laws and prosecute suspects. Here, the international scope of the Internet and other ICT networks can raise some problems. European data protection legislation may easily be avoided when operating from another country without data protection laws. Global harmonisation of data protection laws seems to be imperative for overcoming this problem, but it is to be doubted whether this goal will ever be accomplished.

Personal data protection is currently not functioning properly, particularly in the light of profiling. Both the European Union and several data protection commissioners agree to this and are currently investigating the exact nature of the problem and how it may be solved.[63]

5.2 Intellectual Rights in Profiling Processes

The European Data Protection Directive is a tool to protect the informational privacy of persons or groups. One of its foundations is the principle of transparency according to which the data subject has the right to be informed when his personal data are being processed. This principle finds expression in article 12(a) of the Directive in which the data subject is granted the right to access "knowledge of the logic involved in any automatic processing of data concerning him". In this deliverable this means access to the logic of the profiling software used in financial institutions.

---

[61] In 1996, research surveys by Equifax and Louis Harris & Associates indicated that about 55 per cent of people in the U.S. are privacy "pragmatists", who are willing to trade personal data depending on a number of factors, including the benefits they will receive in return.
See http://www.mindspring.com/~mdeeb/equifax/cc/parchive/svry96/docs/summary.html.

[62] See also Bygrave (2002), p. 311, Sujdak (2001). For some empirical research on privacy perceptions (though in a different context), see Koops, B.J., & Vedder, A.H. (2002). Privacy in Criminal Investigations: A Survey. Criminal Investigation and Privacy: Opinions of Citizens. *The Computer Law and Security Report, 18*(5), 322-326.

[63] European data protection authorities acknowledge the inadequacy of the current directive. For instance, the British Information Commissioner's Office (ICO) recently started research: *UK privacy watchdog spearheads debate on the future of European privacy law*, press release of the British Information Commissioner's Office, 7 July 2008. www.ico.gov.uk/upload/documents/pressreleases/2008/ico_leads_debate_070708.pdf. Also the European Commission has indicated its need for research on different approaches. See *Contract Notice 2008/S 87-117940, regarding the Comparative study on different approaches to new privacy challenges, in particular in the light of technological developments*, published on http://ted.europa.eu.

This right of access will in practice often be blocked by intellectual rights (IR) on the computer programs used. The framers of the Data Protection Directive were aware of this problem. In Recital 41 of the Directive they stated that although the right of access "must not adversely affect trade secrets or intellectual property rights in particular the copyright protecting the software [..] these considerations must not, however, result in the data subject being refused all information." In order to strike an effective balance in this tension between data protection and intellectual rights the applicable regimes of Intellectual Property Rights (IPR) protection have to be analysed. The central question here is: Which legal objects can be identified in profiling processes and which intellectual rights can be vested in them? In order to address this question it is important to closely look at the distinguished steps in the technological process of profiling outlined in section 2.3.2. It turns out that different regimes of legal protection by intellectual rights will be applicable to the flow of information within these systems. These regimes are copyright (or *droit d'auteur*), patent law and trade secrets. Information is thus subject to different legal constraints of protection according to the phase of the process and the kind of legal object it can be classified as. We can distinguish the following relevant legal objects: databases, profiling software, profiles.

5.2.1 Databases

The step in the profiling process in which data are collected and pre-processed often renders a *database*. The legal status has been explicitly tackled in the EU Directive on the legal protection of databases.[64] As a legal object a database is defined as "a collection of independent works, data or other materials arranged in a systematic way or methodical way and individually accessible by electronic or other means".[65] There are two kinds of intellectual rights that can (cumulatively) be vested in a database: copyright and the *sui generis* database right. Copyright can be vested on a database when the "selection or arrangement" of data constitutes the "author's own intellectual creation".[66] These are the general criteria for copyright in collections already protected in article 2(5) of the Berne Convention of 1971. In order to classify as the "author's own intellectual creation" the selection and arrangement of data should bear the imprint of personality of its creator. This means that it cannot be based on just trivial criteria. If these requirements are met the author can exercise four kinds of exclusive rights: the right of temporal or permanent reproduction of the database or parts of it; the right to translate, adapt, arrange or alter the database; the right to distribute the database to the public; the right to communicate, display or perform the database to the public.[67] The latter right means making databases available by other means than by the distribution of copies.[68] This probably includes making databases available "on demand". This protection covers the structure of the database, not its informational contents. This means that on the basis of copyright on databases the contents might be extracted, but that the way they are selected or arranged may not be appropriated.[69]

64 Hereafter: Directive 96/9/EC
65 Art. 1(2) Directive 96/9/EC
66 Art. 2(1) Directive 96/9/EC
67 Art. 5(a-d) Directive 96/9/EC
68 Recital 31 Directive 96/9/EC
69 Hugenholtz, B., Directive 96/9/EC, in Dreier, T., Hugenholtz, B. (eds.), Concise European Copyright Law, Kluwer Law International, Alphen a/d Rijn, 2006.

A second right which can be vested on databases is the *sui generis* right. It applies to databases the making of which required a "substantial investment in either the obtaining, verification or presentation of the contents".[70] There are several differences with the copyright regime of protection. The *sui generis* right doesn't require a measure of creativity, but also protects non-original compilations that have required "the investment of considerable human, technical and financial resources".[71] This includes both qualitative professional skill and quantitative use of labour and money. In this sense the *sui generis* right protects the maker or producer of the database and not its author and is thus a utilitarian right.[72] The maker is "the person who takes the initiative and the risk of investing" and excludes subcontractors.[73] As seen from the definition above, the investment consists of obtaining, verifying or presenting content. This applies to actions in the profiling process of data collection and data preparation. Unlike copyright, this right protects the contents of the database and not its structure. This doesn't mean, however, that it serves as a new right on data or works themselves and thus "doesn't in any way constitute an extension of copyright protection to mere facts or data.[74] The *sui generis* right comprises the two exclusive rights of extraction and reutilization. *Extraction* means the temporary transfer of a substantial part of the database for instance by copying or downloading. *Reutilization* means making available to the public a substantial part of the database.[75]

5.2.2 Profiling Software

The steps of collection, preparation and mining of data can and will often be executed by computer programs.[76] The legal status of *computer software* is complicated. Software can in principle be protected by either copyright, patent or trade secret. These apply to different aspects of computer programs and offer different ranges of protection. In the EU Directive on the legal protection of computer programs[77] it was decided that computer programs should be protected as literary works and thus receive copyright protection. The term computer program is not defined but is supposed to "include programs in any forms, including those which are incorporated in hardware." This protection extends to both source code, object code and assembly code,[78] irrespective of the kind of its physical carrier, such as CD, USB or hardware. In accordance with copyright principles the legal protection extends to the particular expression or form given to the program and not to the underlying ideas and principles.[79] This means that the logic, algorithms and programming languages underlying the software are not protected.[80] As we have seen in the copyright protection on databases, only

70 Art. 7(1) Directive 96/9/EC
71 Recital 7 Directive 96/9/EC
72 Hugenholtz, B., Directive 96/9/EC, in Dreier, T., Hugenholtz, B. (eds.), Concise European Copyright Law, Kluwer Law International, Alphen a/d Rijn, 2006
73 Recital 41 Directive 96/9/EC
74 Recital 45 Directive 96/9/EC
75 Art. 7(2) Directive 96/9/EC
76 In further automated environments of ambient intelligence the KDD stages of interpretation, profile application and decision making will also be executed by autonomic computing software. It remains to be seen how this is affected in the financial sector. On the vision ambient intelligence, see Aarts, E., Marzano, S., The New Everyday. Views on Ambient Intelligence. 010 Publishers, Rotterdam, 2003.
77 Hereafter Directive 91/250/EC
78 Art. 10 TRIPS
79 Art. 1(2) Directive 91/250/EC
80 Recital 14 Directive 91/250/EC

"original" computer programs merit protection, meaning that it is "the author's own intellectual creation".[81] If these requirements are met, the author can exercise three kinds of exclusive rights: 1) "the permanent or temporary reproduction of a computer program by any means and in any form, in part or whole"; 2) the right to translate (from computer language to computer language), adapt, arrange or alter the computer program; 3) the right to distribute the computer program to the public.[82]

Computer programs are thus protected by copyright and, as such, excluded from patentability by article 52 (2c) of the European Patent Convention (ECP) of 1973. In some circumstances, however, they can be considered inventions and thus protected by patents. The exclusion of patentability only applies to computer programs "as such".[83] This is the case when the program is claimed within the framework of a method or an apparatus with technical features and thus constitutes a "computer program product".[84] The execution of the software instructions must be the necessary means to cause a "further technical effect which goes beyond the 'normal' physical interactions between program (software) and computer (hardware)" and must be a means to solve a technical problem. Summarizing: computer languages or codes are considered computer programs as such and receive copyright protection. The technical solution to a technical problem that a computer program may provide is not considered to be the computer program as such, but refers to its function.[85] If it has a technical function or "character" it is patentable as an invention.[86] Meeting these criteria is necessary for patentability, but not sufficient for receiving patent protection. Three further criteria have to be met for this: novelty, inventivity and industrial applicability.[87] A computer program has to be "new" within the available prior art,[88] it has to involve an inventive step that is not obvious to a person skilled in programming[89] and it has to be susceptible to be made or used in any kind of industry.[90] If these requirements are met the inventor acquires the exclusive right to use the profiling software and the possibility to exclude other from this use.

How does this apply to profiling software? The relevant questions here would be whether the profiling software has a technical character, whether it solves a technical problem, and whether it is inventive.[91] The mere fact that the profiling software makes the internal state of a computer hardware change is not in itself sufficient for possessing a technical character. The overall aim of profiling in financial institutions is identification of individuals with respect to fraud prevention or risk assessment These goals will probably not qualify as technical

81 Art. 1(3) Directive 91/250/EC
82 Art. 4(a-c) Directive 91/250/EC
83 Art. 52(3) ECP
84 EPO [Board of Appeal of the European Patent Office], 1 July 1998, T 1173/97
85 The relevant criterion for determining if a program is an invention is not structuralistic, about the "internal physical changes" of the hardware, but functional about "the function performed by the computer program" (EPO, 1 July 1998, T 1173/97).
86 This technical character is already apparent in the requirements for the description of invention in the patent application. Firstly, this description requires the specification of the technical problem the invention deals with and its solution (Rule 27 sub(1)). Secondly, the matter for which the protection is sought has to be defined in terms of technical features (Rule 29(1)).
87 Art. 52(1) ECP
88 Art. 54 ECP
89 Art. 55 ECP
90 Art. 56 ECP
91 The question whether the profiling software is novel exceeds the scope of this deliverable and will be left aside. Furthermore, the industrial applicability is here assumed.

problems. It can also be questioned whether profiling software produces a technical effect which is beyond the normal physical interactions between program and computer. The function of the software is the production of profiles. As such, its function remains mainly on the level of software itself. The (physical) effect of profiling resides in the application of profiles to users. This is not a function of the software itself and thus is beyond the scope of qualification.[92] These considerations make it unlikely that profiling software qualifies for patentability.

Once an inventor has successfully applied for a patent, he or she is required to disclose his or her invention to the public.[93] This may be sufficient reason not to opt for this regime of protection, but to keep the invention secret. One can then opt for the regime of know-how protection. This will only protect the person keeping the secret against several illegal actions of employees (trade secret), competitors (competition law) and other legal subjects. A trade secret offers a company protection by making disclosure by employees an offence under criminal law. One legal form of know-how protection is by trade secret. Several companies protect their computer programs by trade secrets.[94] This is a potential obstacle for the data protection right of access to the logic of profiling.

5.2.3 Profiles

The last and most speculative question pertains to the legal status of the *profile* itself. To our knowledge, no research has been done on this topic yet and no legislation or jurisprudence can give a direct answer to this question. Before we are able to identify as what kind of legal object a profile can be classified we must first have a look at what a profile is. Different perspectives can be taken on profiles. They can be described as knowledge constructs, patterns as the outcome of a process, correlations of data, sets of rules or classificatory categories. The problem is that depending on the description taken on profiles, different IR conclusions could be reached by applying the distinctive criteria of the IR regimes. Some of these qualification ambiguities will be presented here.

As a preliminary clarification, it has to be mentioned that the question about the IR status of profiles only pertains to the construction of profiles, not to their application to people. This already brings us to a first meaning of profiles: profiles can be considered a *knowledge construct* representing a subject. This however is legally too vague to enable IR classification. A second meaning of a profile is a *category* of classification. In copyright (abstract) ideas are excluded from legal protection; only the expressive form of the work receives protection. In the case of profiles, it could be asked whether profiles, viewed as classificatory categories, constitute “ideas”, or are themselves “expressions” of underlying ideas or mathematical concepts. What would be the expressive elements in profiles that would constitute the own

92 This could change when the profiling software is embedded in everyday physical objects together with software for autonomic computing as part of the vision of Ambient Intelligence. In this case the software will produce a further technical effect in these objects. It remains to be seen if and how Ambient Intelligence can be applied in practices of financial profiling.
93 Art. 83 ECP
94 Kamp, M., Weichert, T., Scoringsysteme zur Beurteilung der Kreditwürdigkeit – Chancen und Risiken für Verbraucher, Kiel, 2005.

intellectual creation of the author?[95] We come back to this question when discussing profiles as databases.

A third meaning of a profile is a *correlation of data*. This correlation is the product of profiling software, but is not itself a computer program. Copyright on software is thus not applicable. The profile doesn't constitute an invention either, for the same reason profiling software doesn't constitute an invention: it is not a technical means and it doesn't solve a technical problem. Can these correlations of data be classified as a database?[96] The definition of a database was "a collection of independent works, data or other materials arranged in a systematic way or methodical way and individually accessible by electronic or other means".[97] We can thus ask the question if a correlation of data that is the result of data mining from a database, can itself be classified as a collection of data and thus as a database? In order to answer this question, it is important to distinguish between profiling results and the representation of this output. The output of data mining, the profile, is often a structured decision procedure for assigning new data to an appropriate class. Depending on the function of the algorithm used and the complexity of the results, a certain mode of representation can be chosen. Representation allows an explanation of how the profiles were constructed in a form that is understandable for humans. In case more deterministic algorithms are chosen, possible modes of representation are decision tables, decision trees, sets of classification rules (like Amazon's slogan "people who have bought this item also bought ..." or "if user X satisfies A and B, then he should be classified in class C") or association rules. Some algorithms like neural networks, however, just render a real number output that can just be further computed. The construction process is here a black box.[98] The latter case will not classify as a database. When we describe a profile as a set of rules, at first sight it is unlikely to constitute a database. They rather resemble classical algorithmic steps that are excluded from copyright protection. When we represent the same output by decision tables or decision trees, however, profiles do seem to arrange data in a systematic or methodical way required by the database directive. These contrary results point towards the same ambiguity as discussed above in case of the idea/expression classification.

If we assume for the moment that profiles do constitute databases, we would then have to analyse whether the profile could be protected by copyright. First of all, it must be stated that the fact that something is a personal profile or a group profile, or a distributive or a non-distributive profile doesn't seem relevant since copyright protection is about the form of the database, not about its contents. The nature of the data used (personal or anonymous) is thus irrelevant. The next question is whether "selection or arrangement" of data in a profile constitutes the "author's *own* intellectual creation". To what extent are the profiles produced by the human individual claiming the intellectual rights? This is the case if the profiling software can be seen as a tool used to deliberately create the profile. The resulting profile would then still be determined by the human spirit and therefore the "own" intellectual

95 The question is similar in cases of computer-generated art like creations of fractal Mandelbrot sets. In these cases one can also ask whether these constitute ideas or mathematical formulas or whether the specific input of its creator gives the sets creative expressive elements (Glasser).
96 If a profile doesn't constitute a database, the more general question whether it constitutes a literary or scientific work in the sense of the Berne Convention can still be asked.
97 Art. 1(2) Directive 96/9/EC
98 Anrig, B., Browne, W., Gasson, M., The role of algorithms in profiling, in: Hildebrandt, M., Gutwirth, S., (eds.), Profiling the European Citizen, Cross Disciplinary Perspectives, Springer, Dordrecht, 2008.

creation of the author. But to what extent can the resulting profile reasonably be foreseen and determined by human spirit?

The answers to these questions will depend on the kind of data mining methods used for the construction of the profile and on the fact whether the profiling process is supervised or fully automated. When deterministic algorithms like regression, factor analysis or clustering are used, they will always produce the same result when applied to the same data. In this case, one has to determine whether the form of the resulting profiles is still sufficiently determined by the human spirit of the profiler to constitute his creation. If probabilistic algorithms like neural networks, naïve Bayes classifiers or fuzzy rule induction are used, these results are non-deterministic and unpredictable. In this case, no copyright protection is possible. Whether the created profile can be foreseen or determined by human spirit can also depend on the profiling process being supervised or automated. In the first case, human intervention can occur at several steps in the profiling process.[99] First, "domain experts" have to specify the domain of analysis and the goals of profiling. Second, the expert has to select the data to be used in profiling by choosing the relevant input variables. Third, the expert can select the fields or attributes that are most relevant, instead of using all the collected data. Fourth, the expert has to choose the kind of algorithm to be used for data mining and can intervene in the data mining process itself with heuristics. Fifth, the resulting profiles will have to be interpreted. These interventions show that the skills and cognitive abilities (domain knowledge and bias information) of the profiling expert can thus be of great influence on the outcome.[100] In case deterministic algorithms are chosen, these interventions are probably enough to establish the link between profile and human spirit. Remains to be determined whether these interventions are sufficiently creative to classify as "original". The concepts of creativity and originality have not been uniformly interpreted by courts of the different EU member states to provide a clear answer.

The next question would be whether profiles can be protected under the *sui generis* regime of database protection. This right protected a "substantial investment in either the obtaining, verification or presentation of the contents".[101] If considerable amounts of human, technical or financial resources have been invested in the making of profiles, the investment could be protected by this right. This protection, however, applies to actions of obtaining, verifying or presenting content. "Obtaining" is about gathering the contents or data. "Verification" refers to ensuring the reliability and accuracy of the data and applies to actions of checking, correcting and updating data.[102] These two actions seem to refer to the steps of data collection and data preparation in the profiling process and not to data mining and its results. This would mean that they could not receive *sui generis* protection. However, this will depend on the question whether a profile itself constitutes a "new" database which is separate from the one from which its data were mined and whether, in such cases, the action of mining can be

99 Canhoto, A., Backhouse, J., General description of the process of behavioural profiling, in: Hildebrandt, M., Gutwirth, S., (eds.), Profiling the European Citizen. Cross Disciplinary Perspectives, Springer, Dordrecht, 2008.
100 Canhoto, A., Backhouse, J., General description of the process of behavioural profiling, in: Hildebrandt, M., Gutwirth, S., (eds.), Profiling the European Citizen. Cross Disciplinary Perspectives, Springer, Dordrecht, 2008.
101 Art. 7(1) Directive 96/9/EC
102 Hugenholtz, B., Directive 96/9/EC, in Dreier, T., Hugenholtz, B. (eds.), Concise European Copyright Law, Kluwer Law International, Alphen a/d Rijn, 2006.

classified as "collection". The European Court has neglected the opportunity to give an answer to similar questions.[103]

In the case the answers to these questions are negative, it could very well be that profiles themselves are not subject to any form of protection by intellectual rights. They might only be protected by making them trade secrets. From a data protection perspective this regime of protection would imply the lowest amount of transparency possible for profiles.

5.3 Debate on Credit Scoring in Germany

Up to the present, there has been no specific regulation on profiling and scoring in Germany. The general rules of the German Data Protection Act (Bundesdatenschutzgesetz, hereafter BDSG), which is the enactment of the Directive 95/46/EC in Germany, already apply to scoring. However, much legal uncertainty exists due to diverging interpretations. Currently, the German legislator is working on an amendment of the BDSG addressing in particular the issue of scoring *inter alia* in the financial sector.[104]

5.3.1 Implications of scoring

The possible implications of scoring and profiling techniques for the data subject in general has been object of extensive analysis within FIDIS (Hildebrandt, Backhouse 2005: 73). The German legal situation has been analysed in depth by Kamp and Weichert (2005) already. In this section, the specific problems and implications of the current German practice that led to the current draft bill as well as the core of the planned changes to the BDSG will be discussed.

The market for credit information services and profiling of possible debtors in Germany is dominated by a few major companies which are often aligned with a collaboration of banks or major companies of the telecommunication or mail order business. These companies offer information on debtors to be used in scoring processes by their customers and calculate score values. About 60 companies offer credit information or financial scoring services in Germany. Today many businesses essentially rely on scoring services for credit risk management. Mail order companies need reliable information on the creditworthiness of potential customers. Banks rely on rating as an instrument to meet the requirements of the Basel II framework for the international convergence of capital measurement which requires the lending institutes to provide loss characteristics for granted credits including credits of individuals. At present,

[103] In a request for a preliminary ruling, the European Court of Justice was asked whether the *sui generis* right of art. 7(1) Directive 96/9/EC includes data "derived from the database but which do not have the same systematic or methodical arrangement of and individual accessibility as those to be found in the database", and whether "whenever there is a "substantial change" to the contents of a database, qualifying the resulting database for its own term of protection, the resulting database must be considered to be a new, separate database"? The court did not respond to these questions (ECJ, 9 November 2004, C203/02).

[104] For the draft as discussed here please refer to the version of October 10, 2008, see Bundestagsdrucksache 16/10529, available online: http://dip21.bundestag.de/dip21/btd/16/105/1610529.pdf. This draft is not to be confused with a second bill introducing permission marketing and privacy audits into the BDSG which is debated by the legislative in parallel.

credit scoring raises many privacy issues. The practice in Germany particularly lacks transparency for the data subjects and legal certainty for the processors.

5.3.2 The draft Bill

In this subsection the most relevant planned amendments of the BDSG will be described. The objective of the draft bill is to meet the concerns of data subjects without ignoring the rising need for reliable credit rating within the economy. The benefit for the credit reference agencies, businesses and financial institutions will be a gain in legal certainty. Even though scoring and rating is legal under the present law, some requirements have been under debate and remained unclear. These issues will be regulated and the legal requirements clarified. Data subjects will benefit from an increase in transparency.

Negative payment information
While much personal information, such as the current and past living addresses, the number and kind of credit lines is important for the process, concrete information about the past payment behaviour is frequently used to base a forecast on the future payment behaviour on. This negative payment information consists of hard attributes, such as insolvency of the debtor or oaths of disclosure. Medium and soft attributes consist of order for payment procedures, previous non-payment or delayed payment. Under the current law the requirements for the registration of unpaid debts are debated and unclear. The draft bill tries to find a compromise as it defines these requirements. To ensure that the debtor is aware of the risk that a credit reference agency will be notified about a delayed payment, the debt must be undisputed and the debtor must at least receive two separate demand notes including information about the planned notification. A notification is also permissible when claim is admitted by the debtor or otherwise enforceable by law. In return the draft bill provides a legal foundation for the transmission of personal data by notifying a credit rating agency and waives the obligation to individually consider the data subject's interests.

This solution is heavily debated. The agencies argue that a second demand note would impose additional bureaucratic obstacles and the delay in updating the database might cause damage to other creditors. Data subjects on the other hand gain in transparency as they must be informed about the intention to notify an agency in advance. And as negative payment information may have particularly severe impact on the creditworthiness, it is essential to prevent false notices and to ensure data quality for such information. Data protection authorities criticize the draft. Taking into account that Germany currently experiences that dubious sale promoters massively forge telecommunication contracts or newspaper subscriptions, the new regulation would practically force data subjects to actively react to unjustified claims instead of simply ignoring them to avoid an unauthorised notice to an agency and potentially face financial damage in future.

Notification about financial contracts
New is also the authorisation for commercial banks to notify credit agencies about certain financial contracts (opening a bank account, applying for a loan, granting guarantees for debts of third parties) without the need of an informed consent. This solves potential problems with the requirement of the voluntariness of the consent. At present virtually all banks include a consent clause in their terms and conditions. Due to this lack of alternatives the customers are currently forced to accept the processing. The draft bill clarifies the legal situation by

accrediting the existing legitimate interest of the banks for a defined set of notifications. By waiving the need for an informed consent also the issue of missing voluntariness is solved.

Right of access
The draft bill further intends to provide for more transparency particular by strengthening the right of access. At present the rating agencies tend to refer to trade secrets as an excuse and to actively hinder the right of access which is - at least by the understanding of data protection authorities - already granted by the present law. But even the rating agencies that comply best with the current legal requirements and provide at least proper information about the personal data stored, often refuse to communicate the score itself and won't give details about the process of calculating the score. The agencies argue that disclosing information on the relevant factors and their weighting will enable fraudulent attempts to manipulate the score by providing sugar-coated data, making it difficult for banks to comply with the Basel II requirements.

The draft bill stipulates that the data subjects must be provided upon request with the current scores and those processed within the recent time, the type of data processed and an understandable explanation of the formation of the individual score value. If passed unaltered, the bill will transform the existing legal situation as seen by data protection authorities into the wording of the law. In particular score values are seen as personal data, making the score value object of the right of access and the right to rectification (Kamp and Weichert 2005: 66). However, as the score value is a prediction based on probabilities, the data subjects will still not be able to prove that the value is incorrect; they are therefore limited to rectify the personal data on which the calculation is based.

Information about essential reasons
The draft bill further provides that it will no longer be sufficient in cases of automated decisions to only inform the data subject about the automated processing. Upon request the data subject must be informed also about the essential reasons for the automated decision. The financial institutions criticize that many refusals of credit applications are not based on the score, but on other facts that exclude the data subject from a credit, such as previous breach of contract, being unemployed, under age or insolvent. Instead of providing an explanation of the scoring process, the German association of financial institutions (Zentraler Kreditausschuss, ZKA) suggests that only one core reason for the final decision should be communicated. However accepting this suggestion of the financial institutions would lead to a lack in transparency. According to the Data Protection Directive, processors are already obliged to inform about the essential reasons and the logic of automated decisions to enable the data subjects to verify the accuracy and lawfulness of the processing.[105] This includes due information about major influences on the score value.

Finally, the draft bill provides for the enforcement of the right of access by making non-compliance with the duties an administrative offence that can be prosecuted by the data protection authorities with a fine up to € 300,000. This step is necessary as currently the commissioners lack the power to enforce the right of access in favour of the data subjects.

[105] See recital number 41 of the directive 95/46/EC.

5.3.3 Outlook

If the draft amendment is enacted without major changes, it will provide legal certainty in some questions and provide necessary enhancements for the right of access and its execution. In many points the amendment would not have been necessary, if the credit reference agencies had not denied the existing data subjects rights in the past and thus is clarifying in nature. By the time of the deadline for the editing process (February 2009) the draft bill is still internally debated in the competent ministries. The ratification is planned for 2009.

5.4 The Limitations of Privacy Preserving Data Mining (PPDM)

It is often suggested that the technologies, such as data mining, that create privacy problems should be adjusted and applied in ways that they avoid or minimise these privacy problems. Privacy Preserving Data Mining (PPDM) has been an area of research since the early 1990s. Verykios et al. (2004) described the mining methods that are generally used in the context of PPDM. Oliveira and Zaïane (2004) concluded that there is still no common understanding of PPDM, as different methods for different purposes, e.g. the protection of Personal Identifiable Information (PII) or protection of trade secrets of organisations jointly mining data, are summarised under this term. They also observed a large and rapidly increasing variety of different methods and tools that are available to perform PPDM. These approaches in most cases seem to be limited to one data mining method or even a specialised algorithm. In addition, at that time there was no integrated PPDM solutions available on the market that allows for application independent from the data distribution scheme, the mining method, the algorithm used or even type of attribute (Boolean, numerical etc.) used as basic data.

Based on the results of a search carried out in the context of this research, it can be concluded that these statements are still valid.[106] In addition Oliveira's and Zaïane's (2004) conclusion that further development to achieve integrated solutions including PPDM is necessary is also still valid.

Meints and Möller (2007) investigated how compliance to the Directive 95/46/EC - when applying data mining - can be achieved. Based on literature, a legal analysis and the analysis of a use case (credit scoring), they came to the following conclusions:

"To implement effective privacy protection when applying data mining, it is not sufficient to focus on PPDM methods and algorithms. In addition to this the whole business or governmental process in which data mining is used has to be taken into account. As a result we can conclude that organisational and technical measures taken based on Directive 95/46/EC by applying national data protection legislation in combination with data mining standards allows for quite effective privacy protection. In some cases the strict application of these measures even can make the use of PPDM methods and algorithms obsolete, while in

[106] For an overview on PPDM related literature see e.g. http://www.cs.umbc.edu/~kunliu1/research/privacy_review.html. Further research contributions of 2007 and 2008 can be found when starting an internet search for "Privacy Preserving Data Mining" and the corresponding year.

other cases PPDM potentially can enhance privacy protection further compared to the use of traditional data mining methods. […]"

6 Conclusions

6.1 Answer to the key question

Profiling is very common in the financial sector and due to new technologies it is increasingly used and in more advanced ways. Most financial institutions use profiling for two main purposes. The first reason is for risk management and shaping their core financial business (i.e., assessing risk in order to prevent that costs or payments do not exceed revenues). The second reason is for compliance (i.e., adhering to laws and regulations). These regulations mainly include taking measures against fraud, money laundering and terrorist funding. There is also a set of legal measures (such as Sarbanes-Oxley and Basel II) that is aimed at financial stability. Obviously, these measures are closely related to the risk management purpose of profiling.

Implementing profiling policies, however, raises several problems and does not seem to be effective and efficient, both from a business perspective and from a compliance perspective. From a business perspective, risk assessments have been proven ineffective in the current financial crisis, where unacceptable risks were taken. From a compliance perspective, risk assessments are also ineffective, as it is not too difficult for criminals and terrorists to avoid being noticed during these types of screening.[107] Hence, the key question of this research was: how do financial institutions implement risk profiling strategies from a legal and a informational perspective and with what effects?

From an informational perspective, the risk profiling strategies concern gathering large amounts of data by financial institutions. These data, though not always correct and useful, are usually analysed in order to find risk profiles. Although these risk profiles may be lacking reliability, they are applied to take measures against high risk clients. These high risk clients may be put under close scrutiny, rejected financial services, blacklisted, etc. Clients often have little means of redress as transparency regarding profiling and its implications is lacking.

From a legal perspective, risk profiling strategies usually concern avoiding fines for non-compliance, although some compliance measures may go hand in hand with business goals. For financial institutions it is not always clear how to properly implement legislation. Therefore, they often (try to) discuss strategies with supervisory authorities. The supervisory authorities, however, are often reluctant to provide advice on this, as it may be difficult to apply enforcement afterwards.

The intended effects of risk profiling strategies are to deal with risk properly (including creating financial stability and avoid unacceptable risk) and to deal with crime (including fraud, money laundering and terrorism financing). In general, it can be said that these effects are not achieved with the current profiling strategies that financial institutions use. As mentioned above, in the financial crisis it appeared that unacceptable risks were taken.[108] Furthermore, for criminals and terrorists it is not very difficult to find the gaps in the in

[107] See, for instance, Birrer (2006), Jonas and Harper (2006), National Research Council (2008).
[108] See, for instance, Fratianni and Marchionne (2009). Even before the financial crisis there were indications of poor risk assessments, see, for instance, Wyatt (2002), Kliger and Sarig (2000), Galil (2003).

profiling procedures and avoid being noticed. In the next sections we will describe the main reasons why the current profiling strategies are ineffective from a technological and a legal perspective. These causes of ineffectiveness are the basis for recommendations to improve risk profiling strategies.

6.2 Informational recommendations

From an informational perspective, there are several reasons why the current profiling strategies are ineffective. One of the main reasons is related to the reliability of the strategies. The current profiles often have limited reliability, due to either limited reliability of the data used or limited reliability of the analysis used. When the data is not complete or correct, it cannot be expected that the profiles resulting from it are correct: "garbage in = garbage out". Furthermore, the resulting profiles may be non-distributive (see Section 2.2), which may result in "false positives" and "false negatives". For instance, when a profile is discovered that men in city X, working in pubs and aged 20-25 have an increased risk of committing fraud, this does not mean that all men in this group are actually fraudsters (false positives). Neither does it mean that there are no fraudsters outside this group (false negatives). Collecting a lot of data does not automatically mean that risks are revealed. It would be much better when financial institutions carefully assess which data to collect and use for analysis. Collecting less data may ensure a more targeted approach and, at the same time, causes less security and privacy issues.[109] This requires a thorough analysis of all parties involved regarding their need to know particular data.

Furthermore, when this profile is used to detect risks, its validity is likely to expire. The real fraudsters in this group will be either taken care of by measures of the financial institution (such as rejecting financial services or blacklisting) or try to avoid detection by changing characteristics (such as moving any fraud related activities from pubs to groceries). Hence, over time, the identified group will contain less and less fraudsters. Financial institutions should therefore ensure that their profiles are regularly updated, as their risks constantly change.

As indicated in chapter 4, many financial institutions have set up their risk assessment programs on a large scale and have (partially) automated them. However, they have difficulties identifying the exact scope of these projects, due to the large amounts of customers they have and matching and coupling the different information systems these customers are registered in. Furthermore, they experience difficulties identifying their customers, both natural persons and legal persons. For legal persons, an additional problem is finding the natural persons that are involved in the legal person. Particularly the key decision makers are interesting in this respect, as these people usually determine the risk involved. Persons behind organisations may be difficult to trace, especially when people do not want to be found. Many international companies have parent companies in different countries, causing language barriers, legal barriers and transparency barriers. It may not come as a surprise that many persons that do not want to be traced have accounts in countries with favourable tax climates, strict banking secrecies and lacking transparency.

[109] Since information represents a particular value, collecting and storing more information means more value to protect, as it becomes more attractive for criminals to get access to.

Many profiling programs involve a standardised approach, because of the large number of customers to be screened and profiled. Such an approach is not sufficiently tailored to finding the exceptions and may easily overlook the persons that are suspect. Furthermore, a standardised approach is usually not flexible in quickly changing the strategy, leaving criminals plenty of time to change their strategy without being traced. Financial institutions should focus on more targeted searches. Finding risk, particularly crime, is a cat-and-mouse game where the players are trying to outwit each other. In order to win this game, an ad hoc approach is most suitable, as it provides a creative and flexible approach rather than a generic and predictable approach.

This approach involves more experts doing in-depth investigations on suspects identified by the profiling processes. Rather than the current focus on documents that are easily tampered with, the focus should be on the persons behind the documents and data. For instance, talking to suspects may reveal a completely different picture of this person, easily identifying false positives. In the end, finding risk remains based on human intuition for a significant part. Financial products that nobody really understands contain an inherent risk of not-understanding, let alone a significant risk because the product is not sound. For criminal behaviour this is more or less the same, for instance, when there are patterns that employees of financial institutions with relevant expertise and years of experience cannot explain. It is the human "sanity check" that will help best in avoiding the most serious errors of profiling. It is important to note that it is not recommended to abolish any profiling technologies. However, it is recommended to use them with care.

When errors do occur, however, there should be a possibility of redress for those involved. Currently most financial institutions are not very transparent on their profiling strategies, for both reasons of competitiveness (not revealing risk strategies to competitors) and for reasons of privacy (not revealing any controversial risk strategies to customers). When errors occur, people do not have easy access to file their complaints, have their data removed or changed and have decisions about them made undone. There should me more room for exceptionalities. These rights of data subjects and the openness required for this are discussed in the next section.

6.3 Legal recommendations

From a legal perspective, there are several reasons why the current profiling strategies are ineffective. This is in part due to the laws and regulations that are focused on the process rather than the results. For instance, KYC regulation prescribes identification and profiling of all customers, rather than prescribing sound risk assessment for financial institutions. Basel II is another example, in which in previously 8 % capital margins were prescribed for all situations, rather than high margins for high risks and low margins for low risks. Lawmakers should focus on the intended effects of laws and regulations and get less involved in prescribing how these goals should be achieved. Decisions regarding the procedures and systems to be used for achieving these goals should be taken by the financial institutions together with the supervisory authorities.

When it comes to the unintended effects of the current profiling practices, there are several laws and regulations to protect stakeholders, but many different regimes. Privacy, data

protection and intellectual rights are applicable to profiling practices in financial institutions. These legal regimes have different logics of protection that can be in mutual conflict. Group profiling technologies pose difficulties for the identifiability and data-person-relation criteria in the concept of personal data and thus for the applicability of the regime of data protection. Data quality is of crucial importance in financial profiling due to the major consequences that wrong scores on creditworthiness can have for individuals. The data quality principle, the transparency principle and the individual participation principle tackle these issues. These privacy principles and the rights of notification, access and rectification that are based on them, face several obstacles that complicate their application in practice, such as poly-instantiation of access levels, systems based on a 'guilty-until-proven-innocent' suspicion and non-cooperation by institutions with regard to transparency in data processing.

Some of these problems might be enhanced by a conflicting legal regime that is applicable to financial profiling. The right of access to databases, the profiling programs of credit scoring and the scores or profiles themselves will often be blocked by intellectual rights like copyright, patents, the *sui generis* database right or trade secrets. Each of these objects is subject to different legal regimes which offer varying degrees of protection. In order to be able to judge the legitimacy of such obstruction, the legal status and applicable rights of these informational objects has to be determined. It is a preliminary step that is necessary for striking the balance between data protection and intellectual rights, hinted at in recital 41 of the Data Protection Directive. The balance of this point of friction needs further investigation since it might turn out to be symptomatic for the future interactions between the two main legal regimes that deal with information. Since it is not entirely clear how the current legal regimes protect the interests of stakeholders, particularly data subjects, more clarity is required.

Regarding the privacy and data protection aspects, it can be said that data subjects have difficulties enforcing their legal rights. This is due to several factors. First of all, there is a general lack of transparency. Financial institutions are not transparent about the data on their customers they collect and store. They are not transparent about the ways these data are analysed and used. As a result, people may be confronted with decisions about them, without knowing how these decisions were reached, i.e., on which data and arguments the decisions were based. Thus, people are unaware about what is happening with their personal data. Those data subjects who are being confronted with unwanted decisions about themselves may try to find out what exactly has happened, but financial institutions are not likely to provide detailed explanations about their systems and methods of analysis. This makes it difficult for data subjects to have their data removed or changed or have decisions about them altered. Most data subjects are unaware of their rights in this respect and if they are aware, financial institutions may be unwilling to cooperate. Financial institutions should have clear and open policies on how they collect and process personal data, including statements on which data they collect and for what purposes.

This openness is mandatory from a legal perspective (Articles 10-12 of the EU Data Protection Directive), but in practice there rarely is such openness. Hence the supervisory authorities, in this case the national data protection officers (where possible with the help of the financial sector supervisory authorities), should exercise stronger enforcement. This is important, because, without transparency, data subjects are unable to exercise all their other rights, such as the complaining about lousy security, lousy data quality, excessive gathering of

data, having incorrect data changed or removed and holding financial institutions accountable for non-compliance.

7 Bibliography

Aarts, E. and Marzano, S. (2003) *The New Everyday. Views on Ambient Intelligence*. Rotterdam: 010 Publishers.

Artigas, C. T. (2004) *A Review of Credit Registers and their Use for Basel II*, Financial Stability Institute, Bank for International Settlements.

Basel Committee on Banking Supervision (1999), *Bank's Interaction with Highly Leveraged Institutions*. http://www.new-rules.org/docs/ffdconsultdocs/bcbs1999.pdf

Basel Committee on Banking Supervision (2001) *Customer due diligence for banks*, Basel. http://www.bis.org/publ/bcbs85.pdf

Birrer, F.A.J. (2006) Data mining to combat terrorism and the roots of privacy concerns, *Ethics and Information Technology*, 7, p. 211-220.

Bing, J. (1986) Beyond 1984: the law and information technology in tomorrow's society, *Information Age*, Volume 8, Number 2, p. 85-94.

BSA (1970) http://www.federalreserve.gov/boarddocs/supmanual/bsa/7-00bsaman.pdf

Bygrave, L.A. (2002) *Data protection law; approaching its rationale, logic and limits*, Information law series 10, The Hague/London/New York: Kluwer Law International.

Comptrollers Handbook (2000) *Bank Secrecy Act/Anti-Money Laundering*, Comptroller of the Currency, Administrator of National Banks, U.S. Department of the Treasury. http://www.occ.treas.gov/handbook/bsa.pdf

Custers, B.H.M. (2004) *The Power of Knowledge; Ethical, Legal, and Technological Aspects of Data Mining and Group Profiling in Epidemiology*, Tilburg: Wolf Legal Publishers, pp. 300.

Custers, B.H.M. (2007) *Risk Profiling of Money Laundering and Terrorism Funding; Practical Problems of Current Information Strategies*, Proceedings of the 9th International Conference on Enterprise Information Systems, 12-16 June 2007, Funchal, Portugal.

Denning, D.E. (1988) Lessons learned from modelling a secure multilevel relational database system. In: *Database Security: status and prospects*, C.E. Landwehr (ed.) Elsevier science publishers, IFIP.

Dreier, T., Hugenholtz, B. (2006), *Concise European Copyright Law*, Alphen a/d Rijn: Kluwer Law International.

Ehmann, E. and Helfrich, M. (1999) *EG-Datenschutrichtlinie*, Köln: Otto Schmidt Verlag.

Fayyad, U.M., Piatetsky-Shapiro, G. and Smyth, P. (1996) From Data Mining to Knowledge Discovery: An Overview. In: *Advances in knowledge discovery and data mining*, U.M.

Fayyad G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy (eds.) Menlo Park, California: AAAI Press / The MIT Press.

Fratianni, M.U., and Marchionne, F. (2009) The Role of Banks in the Subprime Financial Crisis. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1383473

Galil, K. (2003). *The quality of corporate credit rating: An empirical investigation*. EFMA 2003 Helsinki Meetings. European Financial Management Association.

Gerding, E. (2008) The Subprime Crisis and the Outsourcing of Financial Regulation to Financial Institution Risk Models: Code, Crash, and Open Source, *Washington Law Review*, forthcoming. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1273467

Hildebrandt, M., and Kindt, E. (2009) *FIDIS Deliverable D7.12: Report on Biometric behavioural profiling (BBP) and Technological Transparency Tools (TETs),* forthcoming Frankfurt a.M.

Hildebrandt, M., and Gutwirth, S., (2008) *Profiling the European Citizen. Cross Disciplinary Perspectives*, Dordrecht: Springer.

Jappelli, T., and Pagano, M. (2002) Information sharing, lending and defaults: Cross-country evidence, *Journal of Banking and Finance*, vol. 26, pp. 2017-2045.

Jentzsch, N. (2007) *Financial Privacy: An International Comparison of Credit Reporting Systems*, 2nd edition, Berlin/Heidelberg: Springer Verlag.

Jentzsch, N. (2007) Do We Need a European Directive for Credit Reporting, *CESifo DICE Report*, no. 2, pp. 48-54.

Jonas, J., and Harper, J. (2006) Effective Counterterrorism and the Limited Role of Predictive Data Mining, *Policy Analysis*, Cato Institute, No. 584, December 11, 2006.

Kamp, M., and Weichert, T. (2005), *Scoringsysteme zur Beurteilung der Kreditwürdigkeit - Chancen und Risiken für Verbraucher*, Kiel.
http://www.bmelv.de/nn_760530/SharedDocs/downloads/02-Verbraucherschutz/Markt/scoring,templateId=raw,property=publicationFile.pdf/scoring.pdf.

Kaufman, G. G., and Scott, K. E. (2003) What Is Systemic Risk, and Do Bank Regulators Retard or Contribute to It? *The Independent Review*, vol. VII, pp. 371-391.

Kliger, D. and Sarig, O. (2000), The Information Value of Bond Ratings, *Journal of Finance*, December: 2879-2902.

Lieber, R. (2009) American Express Kept a (Very) Watchful Eye on Charges, *New York Times*, January 31st 2009.

Meints, M. and Möller, J. (2007) Privacy Preserving Data Mining – a Process Centric View from a European Perspective, *FIDIS Inhouse Journal*, vol. 1, pp. 1-9. http://www.fidis.net/fileadmin/journal/issues/1-2007/Privacy_Preserving_Data_Mining.pdf

National Research Council (2008) *Protecting Individual Privacy in the Struggle Against Terrorists: A Framework for Program Assessment*, Washington DC: The National Academies Press.

Oberlé, V. (2000) *Data Mining: eine Einführung*, University Karlsruhe. http://www.oberle.org/data-mining.pdf

OFAC List (2006) http://www.ustreas.gov/offices/enforcement/ofac/sdn/t11sdn.pdf

Oliveira, S. R. M. and Zaïane, O. R. (2004) Toward Standardization in Privacy-Preserving Data Mining, *Proceedings of the ACM SIGKDD 3rd Workshop on Data Mining Standards (DM-SSP 2004)*, pp. 7-17, Seattle.
http://www.cs.ualberta.ca/~zaiane/postscript/dm-ssp04.pdf

Patriot Act (2006) http://www.epic.org/privacy/terrorism/hr3162.html

Piatetsky-Shapiro. G., and Frawley. W.J. (1998) *Knowledge Discovery in Databases*, Menlo Park, California: AAAI Press/The MIT Press.

Public Interest Research Group (1998) *Mistakes do happen: Credit report errors mean consumers lose*, Public Interest Research Group.

Schweizer, A. (1999) *Data Mining, Data Warehousing*, Zürich: Orelli Füssli Verlag AG,.

Simpson, G.R. (2005) How Top Dutch Bank Plunged Into World of Shadowy Money, *Wall Street Journal*, 30th December 2005, p. A1.

Sujdak, E.J. (2001) *Ethical issues with target marketing on the Internet*, paper presented at the International Symposium on Technology and Society (ISTAS 01), July 6-7, 2001, Stanford, Connecticut.

Treiblmaier, H., et al. (2004) Evaluating Personalization and Customization from an Ethical point of View: an empirical study. In: *Proceedings of the 37th Hawaii International Conference on System Science*, p. 2.

Vedder, A.H. (1999) KDD: The challenge to individualism, In: *Ethics and Information Technology 1*, p. 275-281.

Verykios, V. S. B. E., Fovino, I. N., Provenza, L. P., Saygin, Y. and Theodoridis, Y. (2004) State-of-the-art in Privacy Preserving Data Mining, *SIGMOD Record*, vol. 33, no. 1, pp. 50-57, New York. http://dke.cti.gr/pubs/journals/sigrec03.pdf

Won Kim (2002) Personalization: Definition, Status and challenges ahead, *Journal of Object Technology*, vol. 1, n 1, May – June 2002, p. 31.

World Bank (2003) *Access to Credit Project: Results on the Global Survey of Private Credit Bureaus in 33 Countries; Results on the Global Survey of Public Credit Registers in 64 Countries*.

Wyatt, E. (2002) Credit Agencies Waited Months to Voice Doubt About Enron, *New York Times*, February 8, 2002.

US Sanctions list (2006) http://www.ustreas.gov/offices/enforcement/ofac/programs/

8 Annex 1: Glossary

BDSG	Bundersdatenschutzgesetz
CAID	Chi-squared Automated Detection
CART	Classification and Regression Trees
CRISP-DM	Cross-Industry Standard Process for Data Mining
ECP	European Patent Convention
FI	Financial Institution
FIU	Financial Intelligence Unit
HLI	High Leveraged Institutions
IPR	Intellectual Property Rights
IR	Intellectual Rights
KDD	Knowledge Discovery in Databases
KYC	Know Your Customer
LTCM	Long Term Capital Management
OECD	Organisation for Economic Co-operation and Development
OFAC	Office of Foreign Assets Control
PCR	Public Credit Register
PII	Personal Identifiable Information
PPDM	Privacy Preserving Data Mining
SAR	Suspicious Activity Report
SEPA	Single Euro Payment Area
ZKA	Zentraler Kreditausschuss