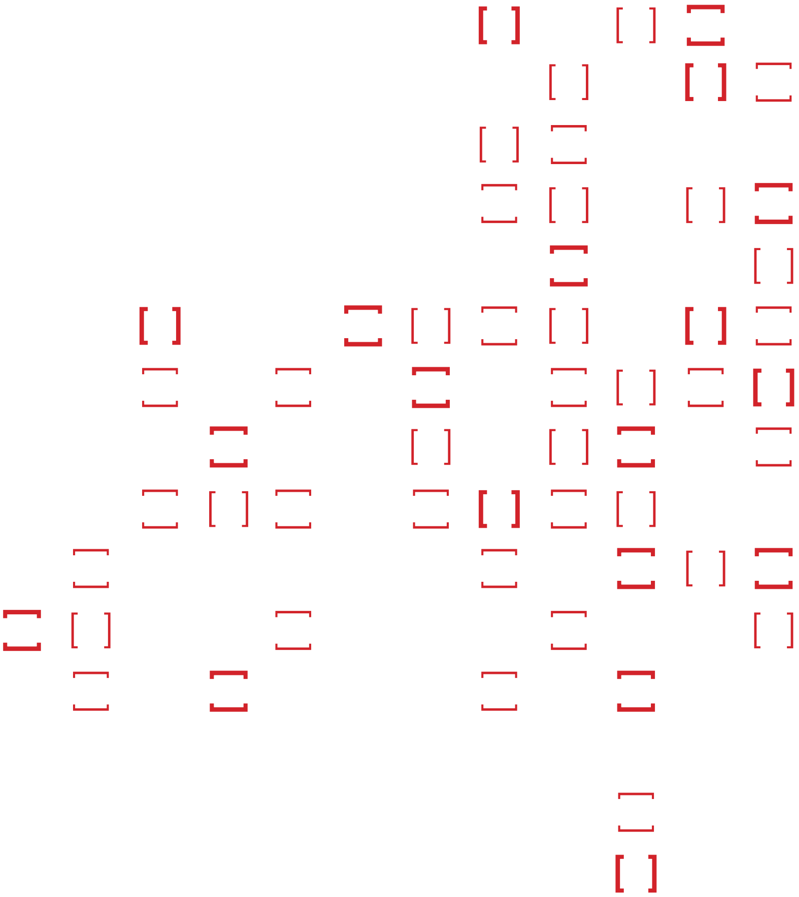
Working Paper Defining Profiling

V. Ferraris AMAPOLA F. Bosco,
G. Cafiero, E. D’Angelo, Y. Suloyeva
UNICRI

prof[L]ng
PROTECTING CITIZENS' RIGHTS FIGHTING ILLICIT PROFILING



prof[l]ng

PROTECTING CITIZENS' RIGHTS FIGHTING ILLICIT PROFILING

With financial support from the “Fundamental Rights
and Citizenship Programme” of the European Union

Table of Contents

Abstract

Profiling is a highly evocative term with multiple meanings, used in both specialist and non-specialist contexts. Drawing attention to the innovative feature of profiling as a form of non-representational, probabilistic knowledge, this paper focuses on machine profiling. It aims to elaborate a suitable definition that captures the main features of this new form of generating and applying knowledge.

The paper is divided in four parts. Part one explores the distinctive elements of profiling. It discusses some existing concepts and distinctions (such as the meaning of organic, human and machine profiling; non-automated and autonomic profiling; group and individual profiling; direct and indirect profiling) and it provides basic information on Knowledge Discovery in Databases and data mining, as key enablers of profiling. It also presents the most relevant sources of profiling, such as behavioural, biometric and location data.

Part two discusses the EU legal framework, including the present discussions on the proposed data protection Regulation and Directive, together with relevant recommendations of the Council of Europe to highlight how profiling is defined and conceptualised in the fields of data protection and anti-discrimination.

Part three gives an overview of different domains of application, including the security, law enforcement and counter-terrorism domain, the financial sector, healthcare, employment, marketing, and social media.

In the final part, the paper develops a definition of profiling. Building on the work of Mireille Hildebrandt, and taking into account insights from the conceptualisation of profiling in other academic literature, law and policy, and from the application areas, the following definitions are proposed of profiling and related concepts.

Profiling is a technique to automatically process personal and non-personal data, aimed at developing predictive knowledge from the data in the form of constructing profiles that can subsequently be applied as a basis for decision-making.

A profile is a set of correlated data that represents a (human or non-human, individual or group) subject. Constructing profiles is the process of discovering unexpected patterns between data in large data sets that can be used to create profiles. Applying profiles is the process of identifying and representing a specific subject or to identify a subject as a member of a specific group or category and taking some form of decision based on this identification and representation.

Keywords: definition, profiling, data mining, EU legal framework, domains of application.

Introduction

Profiling is a highly evocative term with multiple meanings, frequently used in specialist languages and in non-specialist contexts.

For most Profiling is likely to stand for a technique used to discover serial killers. The word is commonly used to identify a technique to classify individuals in order to control and fight against crime and criminals. However, as will be demonstrated in this paper, profiling is much more than that.

Profiling is a new form of knowledge that makes visible patterns “invisible to the naked human eye” (Hildebrandt, 2009c). In particular profiling in its new and most intriguing applications is mainly a form of non-representational knowledge: “profiles do not describe reality, but are detected by the aggregation, mining and cleansing of data. They are based on correlations that cannot be equated with causes or reasons without further inquiry; they are probabilistic knowledge” (Fuster et al., 2010, p.2).

Profiling represents a shift from the idea that knowledge is the result of tested hypothesis. Profiling generates hypothesis: “the correlations as such become the ‘pertinent’ information, triggering questions and suppositions” (Ivi, p.2). The researchers do not need necessarily to know in advance what they are looking for.

The project will focus on what is academically referred to as “machine profiling” (Hildebrandt, 2006; see infra, 1.1). This paper aims to clarify the debate on the definition of machine profiling and generate a common understanding to be used in the Profiling Project.

In order to do this it will progress in four stages. Part one explores the distinctive elements of Profiling identified by scholars; part two discusses the official EU legal Framework together with the recommendations of the Council of Europe in terms of the definition of profiling and data protection; part three gives a brief overview of the different domain of applications. Part four will draw on all the other elements to present a conclusive definition of profiling that will be adopted throughout the Profiling Project.

1. Academic definition

In dictionary, profiling is defined as “the act or process of extrapolating information about a person based on known traits or tendencies, e.g. consumer profiling”; “the act of suspecting or targeting a person on the basis of observed characteristics or behaviour, e.g. racial profiling” (Webster dictionary on line).

Roger Clarke in a 1993 article (Clarke, 1993, p. 1) introduces profiling as a “dataveillance technique”, a “process of creating and using a profile”. Among the few definitions available at that time, Clarke identifies two - referring to two different domains of application - as useful to find a comprehensive one.

The first one (Marx and Reichman, 1984, p. 429) underlines profiling as a method of “systematic data searching” that allows police officers to “correlate a number of distinct data items in order to assess how close a person or a event comes to predetermined characterisation or model infraction”. This definition emerges related to the discovery techniques applied by law enforcement officers.

The second one (Novek et al., 1990) comes from a paper discussing the value of information (customers list, in this case) as a commodity in marketing research.

The paper underlines that “statistical application like regression analysis, non-responder segmentation and models for recency, frequency and monetary values (...) have enabled marketers to ignore unlikely prospects and concentrate on an elite group of potential customers, the lucrative multi-buyers, habitual catalogue and telephone shoppers whose names command the highest prices” (Ivi, p. 529).

These basic and somehow generic definitions underline some common features:

- the central role of data and quantitative techniques;
- categorisation as one of the main characteristic;
- the deduction of new information from something already known (a behaviour, a Specific characteristic, etc.);
- the use of this information for some purposes, i.e. the importance of domains of application.

Clarke (1993, p.2) tried to give a comprehensive definition of profiling, taking into account the different purposes: “profiling is a technique whereby a set of characteristics of a particular class of person is inferred from past experience, and data-holdings are then searched for individuals with a close fit to that set of characteristics”. This definition gives us a overall understanding of what profiling is: profiling is a process of construction of a series of information (a profile), which is then applied to something or someone (individual or group) by techniques of data elaboration.

However in these terms profiling remains a broad concept that warrants a deeper analysis. In the following pages the definition of profiling will be examined by exploring some distinctive key elements of the practice in the specific context of D.P. This exam will allow, together with the exploration of the official definition and the different domains of application, to focus what is the subject of the Profiling Project.

1.1 Relevant distinctions in profiling

Organic, human and machine profiling

In order to better understand the essential nature of profiling, Hildebrandt (2008a) introduces a distinction between organic, human and machine profiling. Organic profiling is a form of profiling carried out by non human organisms, that allows them to know the surrounding environment and consequently to properly adapt for their own survival. The interesting aspect of organic profiling is that the information gathered by the non human organisms (e.g. a plant) is the result of the interplay between non human organisms and environments, meaning that the organism is continuously gathering and processing information. It is a process of continuous updating.

Moreover, the non-human organism does not consciously select information. This procedure is done automatically by the living organism, and it is essential to guarantee its own survival. The reason why organic profiling is helpful in defining profiling is because, as it will appear clear later, there are many aspects in common with the profiling done by the machine.

What does it make human profiling different from profiling by the non human organisms?

Human profiling is different because a human is capable of intentional action and conscious reflection: humans may "consciously reflect upon different courses of action and intentionally prefer one alternative to another" (Hildebrandt, 2008a, p. 27). It is not relevant that "most of human actions are neither intentional nor conscious" because "conscious reflection is the incentive to create new habits which will again move out from the zone of intentional action, but did originate from it" (Ivi). In other words humans act unconsciously or unintentionally because those acts are habits, previously defined by human consciousness.

Machine profiling is more seminal to organic profiling than to human profiling in the sense that it does not imply intentional action or conscious reflection. However, unlike the other types of profiling it is not self-sufficient- any machine needs "an initial software architecture provided by human intervention" (Ivi, p. 28). This disambiguation allows us to understand that profiling is an everyday experience of reducing complexity. Human beings tend to categorise and generalise what happens to them in order to make reality more easily understandable. Living organisms do the same to survive to the surrounding conditions. Machines can be programmed by human beings to automatically process information.

From non-automated to autonomic profiling

The categories of organic, human and machine profiling help to introduce a further and crucial categorisation: non-automated, automated and autonomic profiling. The three categories can be defined as following:

Non-automated profiling is a form of reasoning that does not rely on any process of automatization.

Automated profiling is based on “automated functions that collect and aggregate data” and develop into “automation technologies that can move beyond advice on decision-making, taking a load of low-level and even high-level decisions out of human hands” (Hildebrandt, 2008a, p.28).

Autonomic profiling describes the process whereby the human role is minimized and the decision making process is entirely driven by the machine (Hildebrandt, 2006, 2008a). Autonomic profiling “goes one step further than automated profiling” (Hildebrandt, 2006, p. 550). Ambient Intelligence and Internet of Things are based on autonomic profiling. The machines drive the decision making process, providing for a readjusted environment based on their profiling and without calling for human intervention.

Within the current technological evolution framework, autonomic profiling is not yet the prevalent paradigm; it rather represents a future development of existing forms of automated profiling.

Group and individual profiling

Group profiling identifies and represents a group. The group may consist of a community (i.e. an already existing group) or of a group of people sharing one or more common attributes. The members of the Catholic religion are an example of community, the group of ladies with red hairs and green eyes is not, but it is a group that share those common features.

Group profiling can be classified in distributive group profiling or non-distributive group profiling (Vedder, 1999). A distributive group profile identifies a certain number of people with same attributes. All the members of the group share the same characteristics. For example, the group of the supporters of a soccer team are all identified by being supporters of that team: being supporter is true for the entire group and also for each member of it. In this case the profile can be applied to the group and to any single member because it is also an individual profile.

On the contrary, a non-distributive group profile identifies a certain number of people who do not share all the attributes of the group’s profile. For example, the group of people with higher risks of cardiovascular diseases is profiled by the occurrence of a certain numbers of risk-factors (e.g. specific life-style habits, presence of disease in the family members, stressful conditions at work, etc.). One person may be identified as a member of this group without having the same attributes and without sharing all the attributes. This kind of profiling has a higher probability of mistakenly identify people as members.

Non-distributive group profiles are the most common because it is rare that a large group of people shares all the same attributes. This kind of profiling is always probabilistic and when it is applied there is always a certain degree of risk of inaccuracy. For example there are early prevention programs, which selected participants on the basis of some risk factors (i.e. family composition, school attendance, etc.). Whether the consequent definition of “minor at risk” is based on a non-distributive group profile, the definition would be applied to all minors sharing just some attributes, with a higher possibility of mistakenly include some subjects.

Personalised or individual profiling relies on a set of attributes belonging to a person. It can be used to identify an individual among a group or to infer some of its characteristics.

Direct, indirect profiling

In an attempt to provide a clearer definition of individual and group profiling, Jaquet-Chiffelle (2008) introduces the distinction between direct and indirect profiling. The definition of direct and indirect profiling may apply to both individual and group profiling.

Direct profiling implies that data are collected from one single person or a group and the information derived from the data elaboration will be applied just to the same person or group. It is used to better define that person/group or to deduct future behaviours, habits, etc.

Indirect profiling involves the collection of data from a large population. Individuals are then identified using the attributes emerged from this data collection. The applied profile derived from data referring to other subject. As underlined by Jaquet-Chiffelle indirect individual profiling is what Amazon does each time by suggesting a book to someone based on the purchases done by other people[1]..

Indirect profiling relies on categorisation and generalisation and consequently has a higher degree of uncertainty than direct profiling. As will be described in section 2 on official definition this distinction is relevant from a legal point of view because legislation in general protects personal data but does not provide any legal support in constructing profiles derived from other’s people data or other kind of data.

1.2 Key enablers of profiling

Knowledge Discovery in Databases and Data Mining

Profiling is based on a technique called Knowledge Discovery in Databases (KDD). This technique significantly differs from other data analysis techniques because it provides “its users with answers to questions they did not know to ask” (Zarsky, 2002-2003, p. 6), in brief to discover information hidden in the data.

KDD can also be referred to as Data Mining (DM). Since the terms are used differently in the literature, it is worth underlining that here Data Mining (DM) will be written in the title case format when used as a synonym of Knowledge Discovery in Databases (KDD) to define the entire process. When data mining refers to a specific step of the process (i.e. the application of algorithms), it is written in lower case letters.

[1] Jaquet-Chiffelle (2008), p.63.

The best-known and most widely used definition of Data Mining is the “nontrivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data” (Fayyad, 1996, p.6).

KDD or DM is carried out through several steps (cfr. Hildebrandt 2008a; Zarsky 2002-2003):

1. Recording of data, the operation of collecting or acquiring data;
2. Data warehousing and data cleansing- the preparation of the data (organisation and cleaning) in order to ready it for use;
3. Data mining, application of algorithms by various methods and practices (cfr. Hastie et al, 2009);
4. Examination and interpretation of the results;
5. Follow-up (testing, correcting, etc.);
6. Application of the profiles.

The Data Mining process could achieve both descriptive and predictive aims. Descriptive Data Mining offers a better understanding of the information used in the process. In contrast, Predictive Data Mining generates new information based on the collected one. The analysis aims to “predict outcomes prior to their occurrence” (Zarsky 2011, p. 292).

In both cases, but in particular in the predictive analysis, Data Mining is an automatic procedure. However the role of human beings is still relevant. In step 2, it is a data scientist that chooses how to organise and how to exclude the unreliable information. In step 3, he/she may decide which data mining techniques apply. Finally, in steps 4 and 5 the involvement of the data scientist can differ depending on how the interpretation and test phases are designed. These phases can be totally automatic or not[2]. If they are not totally automatic “the analyst works through the patterns and criteria set forth by the computer algorithms” (Zarsky 2011, p.293).

The main difference between these two options is the capability of the analyst to explain why a specific result was achieved. Obviously, the inclusion of a human analyst in the process is costly and lengthily but provides advantages in terms of accountability and transparency. Finally, the more complex the Data Mining procedure is (due to the high number of database, the complex analysis, etc.) the harder it becomes to have a significant role of human beings in interpreting data.

Big data, Big data analytics

The Economist reported in its 2012 Outlook that the quantity of global digital data expanded from 130 exabytes in 2005 to 1,227 in 2010 and is predicted to rise to 7,910 exabytes in 2015 (Siegele, 2011). The huge amounts of data produced in ever greater quantities, have given rise to another profiling-related concept: Big Data. Alongside the rise of the phenomenon the processed by which it is analysed, Big Data Analytics, has also gained

[2] Zarsky (2011) describes two options: the non-interpretable process, where “human discretion is minimized to setting the parameters for generating predictive algorithms ex ante” and the interpretable one, where the involvement of human actor is greater, pp. 292-293.

popularity. These terms can be seen as roughly synonymous with KDD and Data Mining, but while the latter emphasize the discovery process itself, Big Data Analytics emphasises the basic resource of the process: Big Data.

Big data analytics is the process of examining large amounts of data of a variety of types to uncover hidden patterns, unknown correlations and other useful information. Such information can provide competitive advantages over rival organizations and result in business benefits, such as more effective marketing and increased revenues.

Big data analytics can be done with the software tools commonly used as part of advanced analytics disciplines such as predictive analytics and data mining. But traditional data warehouses may not be able to handle the processing demands posed by big data. As a result, a new class of big data technology has emerged and is being used in many big data analytics environments. There are various companies developing commercial products, but also many Big Data open source efforts, for example HADOOP, Cassandra and Lucene.

As Ann Cavoukian and Jeff Jonas (2012) pointed out, quoting Gantz and Reinsel (2011), "Big Data technologies" describes a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis."

Big Data is not something new but derives from a long evolution of the capabilities of data analysis using computer resources. The much-hyped term has inspired a host of definitions, many of which involve the concepts of massive volume, velocity and variety of information (McAfee and Brynjolfsson, 2012; Jonas, 2012). In other words, what turns data into Big Data is the amount of information, and the speed at which it can be created, collected and analyzed.This huge amount of data and metadata opens the doors to measuring and monitoring people and machines like never before. And by setting clever computer algorithms loose on the data troves, one can predict behaviour of all kinds, such as: shopping, dating and voting. The results, according to technologists and business executives, will be a smarter world, with more efficient companies, better-served consumers and superior decisions guided by data and analysis. The predictive power of Big Data is being explored in fields like public health, economic development and economic forecasting.

Also within the European Union, the importance of Big Data has emerged: the European Commission is funding a 2-year-long Big Data Public Private Forum[3] through their Seventh Framework Program to engage companies, academics and other stakeholders in discussing Big Data issues. The project aims to define a strategy in terms of research and innovation to guide supporting actions from the European Commission in the successful implementation of the Big Data economy.

Specific sources of profiling

Profiling is based on data i.e. "language, mathematical or other symbolic surrogates which are generally agreed upon to represent people, objects, events and concepts" in

[3] See: http://big-project.eu/

order to produce information defined as “the result of modelling, formatting, organising or converting data in a way that increases the level of knowledge for its recipient” (Liebenau and Backhouse, 1990, p. 2). To put it simply, “information” in this context is data arranged in a meaningful way for some perceived purposes.

Almost any data can potentially be used to profile. Although there are several other sources of data currently used for profiling, for the purposes of this paper, two types will be taken into account as examples; these are behavioural profiling and location based profiling.

Behavioural profiling

Behavioural profiling is the study of patterns of behaviour and the consequent grouping of the subjects according to emerged behaviour that emerges. Two different, yet closely related, logic processes underline behavioural profiling; inductive logic and deductive logic. In case of an inductive logic, the processing of data aims to discover patterns in order to explain observed behaviour (e.g. In order to maximise the appeal of a restaurant for lunch time break data may be process to understand which food is preferred by costumers according to the season). In case of a deductive logic, pattern of behaviour are already known and the data processing search for confirmation or negation through other models of behaviours (e.g. the preferred food is known but profiling is used to anticipate changes in the preferences).

Nowadays two of the most common applications of behavioural profiling are: the behavioural profiling of on-line users and biometric behavioural profiling.

Behavioural profiling of on-line users (or Web profiling) is the tracking and tracing of the activities of web users on internet; it is mainly based on the technology of cookies[4]. Cookies are a small text files sent by the server to the client; they are automatically placed on the users’ web browser without any direct visibility. In order to comply with contemporary legal frameworks, websites need to advise the client about the presence of cookies. This said, in many cases web designers construct the website in such a way that if the users do not accept cookies s/he will not be allowed to proceed to the rest of the content.

Cookies allow content providers to trace, store and use the preferences of web-users, from language settings, to frequent e-mail recipients to purchase goods history.

Today, e-commerce sites are the most relevant context where behavioural profiling is used. For example, whoever has purchased any items through Amazon is aware of the “shopping hints system” developed by Amazon. The systems provides hints for further purchases based on a comparative model that relates the searches and purchases of the individual to those of others that have displayed a similar purchasing behaviour.

Like Amazon, most of the e-commerce websites use profiling to customise their offers to the clients.

[4] Monitoring of IP addresses, using of Javascripts, identification of browser fingerprints are all techniques for web-tracking.

Web-users profiling is not only a prerogative of any single website. The more information is available, the more potential profiling has. The concentration of information in the hands of few powerful websites, i.e. private companies, allows them to profile in such a way that none else can do. This is the case of Google, not only because it is one of the main third-party aggregators and tracks users across websites (Krishnamurthy and Wills, 2009 and Castellucia, 2012) but also because it owns of one the most popular search engines worldwide and one of the most popular email systems- Gmail.

Another example is Facebook which due to its connection with other websites, collects a huge amount of information (see Roosendaal, 2012).

Biometric behavioural profiling

Biometrics refers to "systems that use measurable, physical or physiological characteristics or personal behaviour traits" (T-PD 2005, p. 4). When biometrics data are used in profiling there is a set of technologies named behavioural biometrics. These technologies do not rely on physical features, but measure human characteristics related to conscious and unconscious behaviour.

Behavioural biometrics allow in primis the construction of individual profiles. Profiles are the result of:

- comparison of a presented biometric sample with the biometric data pertaining to one single person (e.g. deduction of gender from person's voice);
- matching of the sample not only with the data of the same person, but also with the biometric data of other data subjects in the same or in others databases.

Behavioural biometrics "can provide useful profiling information such as measure of a person's preferences or mood" (Yannopolus et al, 2008, p.109). It requires high-quality technology (cameras, sensor of different kind, etc.).

The technologies that enable behavioural biometrics are growing in numbers and in quality. They allow emotion and gesture recognition, human gait, voice and signature analysis, keystroke dynamics and mouse movements. The more they increase their performance, the more they produce refined results allowing large-scale application.

Moreover biometric systems can also benefit of research done in other domains of application. Take as example the research based on "participatory sensing approach", meaning that data will be collected also by citizens through mobile apps and sensor devices that they will be wearing. For example, it is possible to apply this approach at the research on air pollution for the benefits of the citizens and, in this case, it will certainly also provide benefits for other domains of application highly sensitive to outdoor condition, such as biometrics.

Location based profiling

Location based profiling works thanks to location based-services (LBS), that are services able to locate someone or something (people, vehicles, potentially any kind of mobile objects) in the territory.

Location or mobility data are available due to the wireless and mobile communication technologies: mobile phones for people[5], GPS for vehicles, RIFD-tags for objects or animals are all tools adequate for this purpose.

Location data do not only provide trivial information on where people and things are in a specific moment in time; they can be processed at a single point in time and space or within a time window or a space area. As the table below shows, the amount of information that can be deduced is massive and in many cases they can be sensitive personal information.

		Spatial dimension	
		At one point	Within an area
Temporal dimension	At one moment	Know about the status quo of time and space at one moment (e.g. a hospital visit)	For individuals: Makes no sense as one can only be in one place at one moment in time. For groups: can reveal relationships, social circles, collaboration.
	Within a time window	Can reveal workplace, home, social context and information about personal preferences (e.g. restaurant type).	Reveals shopping habits, dating habits, driving speeds and other information.

Source: Fritsch, 2008, p.171

On the other hand mobility data also represent an asset for the development of sustainable mobility, i.e. they help to plan public transportations, traffic; to forecast traffic-related problems, etc. (Giannotti, Pedreschi, 2008).

Behavioural profiling and location based profiling make clear which kind of data can be used for automated profiling and how they have a twofold nature: they represent a great potentiality and a great risk.

So far, we have explored the main definitions of profiling within the academic literature. We went through the relevant distinctions within the profiling techniques; we analyzed its key enablers and presented its main sources, by exploring both biometric and location based profiling. In order to complete the landscape of profiling definitions, we will now go through the existing legislation in this field as well as the new proposed General Data Protection Regulation.

[5] It is worth underlining that many mobile phone applications demand the geo-localisation in order to work properly. This is self-evident for some applications (e.g. the ones used to locate services, such as restaurants, pharmacies, etc.). Friedland and Sommer (2010) found out that many people are unaware that photos and videos taken with smart phones or cameras include geo-location information.

2. Official definition

2.1 Current data protection legislation: Article15 of Directive 95/46/EC

It is worth pointing out that the protection of personal data is recognized as a fundamental right in various European and international legal instruments. An important component of European Union legal framework, which regulates the processing of personal data is Directive 95/46/EC of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (hereinafter, the Data Protection Directive, DPD). This instrument has a direct effect in EU Member States because of the legal obligation to implement it into the national legislation.

Even if profiling ad hoc measures are not foreseen, Article 15 of the DPD is of particular relevance. The provisions of this article concern 'automated individual decisions' and thus are closely related to profiling. According to article 15(1): "every person has the right 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." At the same time, article 15 (2) states an exception: "a person may nevertheless be subjected to an automated individual decision if that decision is taken: (a) in the course of the entering into or performance of a contract, provided the request for the entering into or the performance of the contract, lodged by the data subject, has been satisfied or that there are suitable measures to safeguard his legitimate interests, such as arrangements allowing him to put his point of view; or (b) is authorized by a law which also lays down measures to safeguard the data subject's legitimate interests".

Though the term Profiling is not mentioned in this article, the original proposal included the word profile stating that data subjects have the right "not to be subject to an administrative or private decision involving an assessment of his conduct which has as its sole basis the automatic processing of personal data defining his profile or personality"[6]. The essential question is whether the involvement of a human being is always required when it concerns decisions that affect an individual.

In particular this means that a decision can be taken based on a profile, even when this profile is created by automated means only, but the involvement of a natural person in actually taking the decision is required. In the light of Article 15 of the DPD, it is relevant whether the processing is meant to reveal a certain aspect of the personality of an individual

[6] See Art. 14 (2) of the Proposal for a Council Directive concerning the protection of individuals in relation to the processing of personal data (COM (90) 314 final – SYN 287, 13.9.1990 (granting a person the right "not to be subject to an administrative or private decision involving an assessment of his conduct which has as its sole basis the automatic processing of personal data defining his profile or personality"); and Art. 16(1) of the Amended Proposal for a Council Directive on the protection of individuals with regard to the processing of personal data and on the free movement of such data (COM(92) 422 final – SYN 287, 15.10.1992 (granting a right to every person "not to be subjected to an administrative or private decision adversely affecting him which is based solely on automatic processing defining a personality profile").

on which a decision can be based. This situation implies that, first of all, personal data are at stake in the processing, and also that coming decisions will be, based on a "digital persona" (Clarke, 1994)[7]. However, regardless of whether the data contain personal data, the decision will be connected to an individual, thereby constituting the identifiability, which is necessary to speak of personal data. Thus, also the combination with personal data afterwards makes the DPD applicable to the processing of non personal data (Roosendaal, 2010, 2013).

In summary, article 15 does not take the form of a direct prohibition on a particular type of decision-making; rather, it directs each EU Member State to confer on persons a right to prevent them from being subjected to purely automated decisions in general (Bygrave, 2002, p.3).

2.2 The new Directive and Regulation proposals

Given its social and technological context, Directive 95/46/EC did not manage to fully achieve its internal market policy objective, nor to remove differences in the level of data protection actually afforded by the Member States. Since the Directive does not provide for sufficient protection in a fast-developing information society and globalised world, the increasing issues relating to data surveillance and uses call for a new legal framework for the protection of personal data in the EU. In response to these issues, the European Commission released a draft General Data Protection Regulation (GDPR) in January 2012, a set of proposed reforms to the existing EU data protection law. If approved, the GDPR would unify data protection law across all 27 EU Member States.

It is important to point out that article 20 of the GDPR covers the definition of profiling and establishes measures concerned. Accordingly, it gives every 'natural person' "the right not to be subject to a measure which produces legal effects concerning this natural person or significantly affects this natural person," based on automated processing for profiling purposes. In this article the profiling has to be meant to "analyse or predict in particular the natural person's performance at work, creditworthiness, economic situation, location, health, personal preferences, reliability or behaviour" (Article 20, GDPR).

European Commission also released a draft Directive proposal to replace the existing Council Framework Decision 2008/977/JHA of 27 November 2008 on the protection of personal data processed in the framework of police and judicial cooperation in criminal matters. Law enforcement has been identified as an area where improvement was needed, and the previous directive does not provide a comprehensive framework of data protection by law enforcement and judicial authorities in criminal matters.

[7] Roger Clarke (1994) defines in his paper 'digital persona' a model of the individual established through the collection, storage and analysis of data about that person (See WP 2).

Response of the European Parliament

The European Parliament strongly supports the reform proposal, in particular, the replacing of the current Data Protection Directive with a directly applicable Regulation. The Parliament supports in principle a general ban introduced on profiling, as it should be only permissible in limited situations such as with an individual's consent. However, it is of the view that a clear definition on profiling was missing in the reform proposal, so it should be further clarified. Moreover, any such definition should be in line with the Council of Europe Recommendation CM/Rec (2010) 13 (see infra, section iii). In response to the proposal, Jan Philipp Albrecht (Rapporteur for the LIBE Committee, which is leading the European Parliament's position on this matter) published a proposed revised draft Regulation (Draft report on GDPR, 2012). Considering that 'profiling' was covered by the Commission's proposal but not defined, the Draft Report introduces a wide definition. Profiling is defined as "any form of automated processing of personal data intended to evaluate certain personal aspects relating to a natural person or to analyse or predict in particular that natural person's performance at work, economic situation, location, health, personal preferences, reliability or behaviour" (Out-Law, 2012).

Response of the European Data Protection Supervisor

The European Data Protection Supervisor (EDPS) Peter Hustinx announced in his reaction that the proposed Regulation constitutes a huge step forward for data protection in Europe. He supports the main objective of the proposed Regulation to harmonize and simplify the application of data protection principles across the EU. He is convinced that in a technological environment where data processing is rarely limited to territorial boundaries, this will enhance legal certainty both for individuals and data controllers. The EDPS supports the clarification provided by the Proposal on its scope of application and the development of the list of definitions (EDPS, 2012).

With regard to the measures based on profiling, the EDPS supports the provisions of article 20 of the proposed Regulation. The article builds upon the existing Article 15 of Directive 95/46/EC on automated individual decisions, and extends its scope to all types of measures which produce legal effects on a natural person, not only to decisions.

Firstly, the positive aspect is that it would apply not only to processing intended to evaluate certain personal aspects but also to those activities carried out to analyse or predict these aspects, therefore encompassing a broader category of processing. Secondly, it introduces a number of categories of personal aspects, which would fall under the scope of this provision, such as processing concerning an individual's economic situation, location, health and personal preferences.

Thirdly, article 20(2) sets forth the conditions under which this type of processing may take place by way of derogation and it provides data subjects with the right to have human intervention but not with the right to submit their point of view, as is currently provided for in Article 15 of Directive 95/46/EC.

Response of the Article 29 WP

The Article 29 Working Party was set up under the Directive 95/46/EC of 24 October 1995, from which belongs its name. It has advisory status and acts independently and it is made up of a representative of the supervisory authority designated by each EU country, a representative of the authority from the EU institutions and a representative from the EU Commission[8].

As stated in its Opinion 01/2012 on the GDPR[9], the Working Party believes that more must be done to explain and mitigate the various risks related to the profiling.

Last May 2013, the Working Party adopted an advice paper on this issue, were it proposed some essential elements for a definition and a provision on profiling within the new EU legal framework on data protection. In fact, in the light of the increasing usage of profiling technologies in the private and in the public sector and their possible impacts on the basic right to data protection, the Article 29 Working Party deems it is necessary to include a definition of profiling in Article 4 of the General Data Protection Regulation.

Based on the 2010 Council of Europe Recommendation on profiling and the Commission's wording in Article 20(1), the Working Party proposes the following definition: "Profiling" means any form of automated processing of personal data, intended to analyse or predict the personality or certain personal aspects relating to a natural person, in particular the analysis and prediction of the person's health, economic situation, performance at work, personal preferences or interests, reliability or behaviour, location or movements (Article 29 Working Party, 2013).

Within the advice paper, the Working Party also provides suggestions on how to improve the Article 20 of the GPDR.

Regarding the "scope", one of the main limits of article 20 is that it merely focuses on the outcome of profiling rather than on profiling as such. In this sense, a necessary step would be to broaden the scope of Article 20 covering processing of personal data for the purpose of profiling or measures based on profiling, in order to obtain more legal certainty and more protection for the individuals.

To this aim, the following additional elements should be included in the Article:

1) Greater transparency and more individual control on the decision for data subjects on whether or not own personal data may be processed for the purpose of profiling of measures based on it. The Working Party underlines in particular the importance of explicit consent as a legal basis for data processing also in the context of profiling.

2) More responsibility and accountability of data controllers with respect to the usage of profiling techniques.

[8] See: http://ec.europa.eu/justice/data-protection/article-29/

[9] Fore more in depth information check the Article 20 Data Protection Working Party (2012), Opinion 01/2012 on the data protection reform proposal, adopted on 23 March 2012. Available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2012/wp191_en.pdf

3) Need to take a balanced view: the new Regulation should provide for clear rules on the lawfulness and on the conditions for the processing of personal data in the context of profiling [while] leave a reasonable degree of discretion to assess the actual effects and the degree of intrusiveness. In this view, article 20 only applies when profiling does not significantly affect the interests, rights or freedoms of the data subject. However, a mechanism is needed in order to interpret the term "significantly affects" and this task could be best performed by the European Data Protection Board.

2.3 Council of Europe Recommendation CM/Rec (2010) 13

In terms of the definition of profiling, and the debates surrounding it, a particularly comprehensive outlook is presented in the Council of Europe's 1099th Meeting of the Ministers' Deputies (CM/Rec (2010)13). As in the case of most commentators, the Council underlines that the fundamental issue with profiling lays at intersection of the individual's rights, protected by Article 8 of the ECHR, and the vast benefits that society can derive from mining big data. Having outlined the dangers that may arise out of attempting to define something, which is at the interface of public interest and private rights, the members of the Council, draft an outline of the profiling process. They offer a definition that divides the process in three stages: an observation stage, which can also be referred to as the data-warehouse stage, where data from a variety of different sources is collected and stored; a stage where this da ta is analysed, which can also be referred to as the data mining stage; and, a final implementation stage (paragraph 38).

The implementation stage, the Council argues, is what should distinguish the legal definition of profiling from the common term use of the word. Since selecting individuals on the basis of their real characteristics does not constitute profiling, in the Council's opinion it is important to distinguish profiling techniques from other aids to decision-making. In this regard, such a formulation of the profiling definition is very similar to the definition of the technique profiling is based on (Data Mining), mentioned in the section 1, which divides the process in six stages[10]. The Council of Europe offers the example of banking activities. Banks commonly use the term "rich customer profile" to highlight those customers who earn over a certain amount of money per month and have an estate worth over a given amount. Whilst a 'rich customer profile', is indeed a type of 'profile', the Council suggests it should be distinguished from 'profiling' because in order to attribute the title ('rich customer profile') banks rely on solid evidence about that person's credit history rather than inferring things about it from the analysis of other people's behaviour (paragraph 41). On the other hand, when a bank is trying to determine whether or not it should give a loan to an individual, the bank asks the individual to provide a series of answers to seemingly neutral questions. Banks then use the psychometric tools derived from the analysis of behavioural data to determine whether or not to undertake the risk of lending the money. In this case, the council notes that the bank is 'profiling' because it isn't using only factual data pertaining to

[10] The Observation stage corresponds to stage 1 and 2 (recording of data, data warehousing and data cleansing), the data mining stage corresponds to stage 3 (data mining) and finally the implementation stage refers to stages 4, 5 and 6 (examination and interpretation of the results; follow-up (testing, correcting, etc.) and application of the profiles)

individual, but it is inferring characteristics about the individual through the use of statistical data which by its nature can only ever be partially accurate (paragraph 43).

2.4 Existing official definitions of profiling

It is now clear that there have been different approaches to define profiling in the European context. One approach originates from a data protection perspective, while the second one emerges from an anti-discrimination perspective.

Two official definitions from the data protection perspective derive from the Council of Europe Opinion on profiling and the Draft General Data Protection Regulation (GDPR). According to the CoE Recommendation, Profiling – is an “automatic data processing technique that consists of applying a ‘profile’ to an individual, particularly in order to take decisions concerning her or him or for analyzing or predicting her or his personal preferences, behaviors and attitudes” (paragraph 1, CM/Rec (2010)13).

Another official definition comes from the Draft GDPR, where Profiling is defined as “automated processing intended to evaluate certain personal aspects relating to this natural person or to analyze or predict in particular the natural person’s performance at work, economic situation, location, health, personal preferences, reliability or behaviour” (See article 20, Proposal for GDPR, 2012).

As mentioned, the Article 29 Working Party - based on the 2010 Council of Europe Recommendation on profiling and the Commission’s wording in Article 20(1) - proposes the following definition: “Profiling” means any form of automated processing of personal data, intended to analyse or predict the personality or certain personal aspects relating to a natural person, in particular the analysis and prediction of the person’s health, economic situation, performance at work, personal preferences or interests, reliability or behaviour, location or movements”. These definitions describe profiling as a neutral process, while the definitions grounded on an anti-discrimination perspective are concentrated on ethnic profiling.

The European Commission against Racism and Intolerance (ECRI) in its General policy recommendation defines ethnic profiling as “the use by the police, with no objective and reasonable justification, of grounds such as race, colour, language, religion, nationality or ethnic origin, in control, surveillance or investigation activities” (See paragraph 1, ECRI General policy recommendation No 11, 2007). Considering that the racial profiling constitutes a specific form of discrimination, the definition of the above mentioned racial profiling derives from the definition of racial discrimination, used by the ECRI in its General Policy Recommendation N°7 (paragraph 1, ECRI General policy recommendation No 7, 2002); and also from the definition of discrimination used by the European Court of Human Rights in its case law.

However, this last definition is not particularly focused on automated profiling: it is a wider and generic definition, which entails almost every form of profiling.

3. The Purpose of Profiling and the Domain of Applications

3.1 The purposes of profiling

Although technology is never inherently good or bad, its application and effects are not impartial (Kranzberg, 1986).

Profiling is both a practice and a technique: “a specific way of doing things, within specific contexts, and with specific purposes” (Hildebrandt, 2005, p.51). As highlighted by Hildebrandt and Gutwirth (2008), nowadays we are facing an ever- expanding mass of information and we need to understand how we should act on it and what this information means. In this context profiling practices are some of the technologies with more potential to create order in the turmoil of proliferating data.

In order to understand the purpose of profiling, it is important to understand the meaning of profiling (see sections 1 and 2). The effects and the purposes of profiling are determined by the users of the profiling technologies. Thus “the purposes [of profiling] are the explicit objectives, formulated by the data controllers” (Hildebrandt, 2005, p. 52). The effects, on the other hand are the more or less intentional consequences on the profiled subjects.

Profiling techniques can be applied in different contexts, and for different purposes. For instance profiling can be used to: detect potential terrorists or criminals, discover potentially fraudulent or productive employees, and to source new customers in different areas. The purposes are related to the different fields of application, which entail the risks and the opportunities behind the data mining process. “[…] The purpose of profiling practices should be taken into account, as this determines both the adequacy of the construction of profiles and their impact on our world” (Hildebrandt, 2005, p.9).

Overall the purpose of profiling is that of creating order and producing new knowledge from existing data. Even if it can be applied within different application fields, either for security or commercial purposes, usually profiling seeks to predict future behaviour “by relying on the stereotypes learned during the data mining step, classifying the individuals as potential risks or commercial windfalls” (USI, Trilateral, CCSC, 2012, p.30).

3.2 Different domains of application

Profiling practices are increasingly used both in public and private contexts, for different purposes. “The rationales and internal balances discussed in the governmental context cannot be applied directly to the private sector. With private firms, competitive forces […] might play an important role in achieving some of the needed objectives”. The obligations and motivations of governmental entities are different from their commercial counterparts (Zarsky, 2011, p.5).

The distinction between the data processing activities of private and that of public agents is not clearly defined (Jeandesboz, Bigo, Frost, 2011). In some cases public bodies and governmental agencies can tap into personal data held by private organizations for security reasons. For example, the pattern recognition and prediction practices, typical of the commercial practices of dataveillance, "are equally present in data processing schemes setup for policing purposes" (Ivi, p.14). This data processing practice is becoming a characteristic of several criminal justice systems, through the promotion of the so called "intelligence-led policing".

Within the previous sections, some distinctive elements of the definition of profiling have been explored. On the basis of these, the different application domains of this technique will be now investigated.

Firstly, the use of profiling in the law enforcement agencies and security fields will be analyzed, as well as its role in the border controls and in the fight against terrorism. After that, the application of profiling in the financial and health domain, in the employment and marketing sector, and in the world of the social media will be presented and discussed.

Security and criminal investigations domain

Intelligence, security, law enforcement

"Data mining has captured the imagination as a tool that can potentially close the intelligence gap constantly deepening between governments and their new targets-individuals posing a risk to security and the public's wellbeing. Data mining is also generating interest in other governmental contexts, such as law enforcement and policing" (Zarsky, 2011, p.287). Illustrative is the fact that the "Intelligence-Led Policing" is generating increasing interest at government level. This type of policing can be defined as "a strategic, future-oriented and targeted approach to crime control, focusing upon the identification, analysis and management of persisting and developing problems or risks … rather than on the reactive investigation and detection of individual crimes" (Maguire, 2000, in Van Brakel, De Hert, 2011, p.168). It tries to build up intelligence through all kinds of data collection strategies. In particular, profiling techniques bring together pre-emptive and surveillance policing.

The law enforcement agencies are trying to pre-empt crime, instead of merely reacting to the events. In this framework, emphasis is put on "pro-active" and "pre-emptive" policing (USI, Trilateral, CCSC, 2012), which nowadays are core components of internal security activities in many European Member States, as widely recognized by strategic documents - such as the Stockholm Program[11].

Looking at the pre-emptive profiling techniques, the main purpose is to cluster data in order to infer information and thus propose predictions. The profiles obtained do not describe

[11] Stockholm Program is an act providing a roadmap for UE work in the area of justice, freedom and security for the period 2010-2014. See: http://europa.eu/legislation_summaries/human_rights/fundamental_rights_within_european_union/jl0034_en.htm

reality, “but are detected in databases by the aggregation, mining and cleansing of data. The profiles are based on correlations […]” (Gutwirth, Hildebrandt, 2008, p.32).

One example of a “pre-crime database” is the E-CAF system, introduced in England to increase effectiveness in crime prevention and child protection. Several public services have access to this database, including police, social services, schools etc. The idea behind it is to have a way to predict which minors will commit a crime and to intervene before this happens. This system works on the basis of profiling and risk factors for delinquency, which have been suggested in crime prevention-related research. The E-CAF system even makes a decision about the need for further intervention on the basis of the risk score that is assigned to a certain minor (Van Brakel, De Hert, 2011).

The FBI biometric database is a prime example of the police’s use of profiling. According to the FBI, this is the “largest biometric database in the world”, which contains records for more than a hundred million people. Moreover, they are planning the so-called Next Generation Identification, a massive upgrade that “will hold iris scans, photos searchable with face recognition technology, palm prints, and measures of gait and voice recordings alongside records of fingerprints, scars, and tattoos” (Ganeva, 2012, p.1).

Another interesting aspect of the application of profiling in this field is open-source intelligence. The US military defines “Open Source Intelligence” (OSINT) as “relevant information derived from systematic collection, processing and analysis of publicly available information in response to intelligence requirements” (Hayes, 2010, p.1). Thanks to the profiling techniques, such as data mining, disparate pieces of raw OSINT data can be put into the right context and provide knowledge and further pieces of information useful for investigation purposes. For example law enforcement Departments which have an OSINT unit, as NY Police Department, Scotland Yard, Royal Canadian Mounted Police, apply OSINT to the prediction, prevention, investigation, and prosecution of criminals including terrorists.

From all this, it can be derived that Profiling within law enforcement is understood in relation to both pro-activity – that is following electronic traces left by persons/groups targeted by surveillance – and prevention – aimed at assessing a future threat and prevent the event from happening. Law enforcement gathers intelligence, including personal information, and analyzes it in order to allocate police resources accordingly. The practice of Profiling can become problematic “when it is expected to act upon the future, support actions against persons/groups in the name of behaviours they are expected to have” (Jeandesboz, Bigo, Frost, 2011, p.15).

Border control

Another field of application of profiling related to security is the risk assessment during border controls. The increasing flow of passengers through airports, and phenomenon such as the free movement of people in the Schengen area or the attacks of September, 11 in 2001 in the U.S., caused a re-think about how borders can be best protected.

A practical example in the field of border controls is the Passenger Name Record (PNR) System. PNR was developed with customer service purposes, and contains several fields of information ranging from travel-related information to very personal (e.g., dietary requirements) and relational (e.g., travel companions) data. "PNR contains information that can be analyzed and data-mined in conjunction with other intelligence information to identify allegedly high-risk travellers who are either singled out for extra attention at check-in or upon arrival at their destination, or are deemed too dangerous to fly" (International Civil Liberties Monitoring Group, 2010, p.12). Governments consider PNR risk analysis – along with biometrics – as key elements of border management procedures.

The European Council has expressed interest in developing a European passenger name records (PNR) system for law enforcement purposes and in 2008 started to work on a Framework Decision on this matter. With the entry into force of the Lisbon Treaty in 2009, this draft framework, which had not been adopted by the Council by that date, became obsolete (EU, PNR FAQ, 2011). A new draft Directive on the use of airline Passenger Name Record data was proposed in 2011, which is currently being discussed; in April 2013, the European Parliament's Civil Liberties Committee rejected the proposal[12].

Besides the regulation on PNR existing within the EU, three consecutive Agreements between the EU and the US have been drafted in order to try to regulate the existing divergence on this issue, specifically with regard to the related protection of individual privacy. "PNR data obviously constitute 'personal data' within the meaning of EU law […] PNR data transfers system leads to the creation of a database with comprehensive information on all basic individual data, such as residence and workplace, payment preferences, age, etc. Moreover, by collecting and correlating information like 'special meal requirements' or 'seating particularities' or even by 'efficient' use (profiling) of the same individual's name, inferences may be made about such sensitive issues as the religion or health condition of the passengers" (Papakostantinou, de Hert, 2009, p.887).

In conclusion, if on one side the success of profiling practices in the private sector to make distinctions between reliable and unreliable customers can be seen as a good practice and the way forward in the issue of border protection - such as in the case of PNR System - on the other side concerns can be raised with regard to the respect for individual freedom and anti-discrimination –as in the case of racial profiling.

Moreover, another aspect to be considered is that the aforementioned EU-US Agreements tried to legalize a situation created after 9/11, when the US Bureau of Border and Customs Protection (CBP) started asking international air carriers for access to their passenger data, as a measure to counter terrorists[13.]

This leads our analysis to the next domain of application, the role of profiling in the fight against international terrorism.

[12]http://www.alde.eu/press/press-and-release-news/press-release/article/no-eu-pnr-before-data-protection-rules-are-firmly-in-place-41026/

[13] For an in-depth analysis of the PNR Agreements between EU and US look at Papakostantinou, De Hert, (2009), De Hert, Bellanova (2011).

Counter-terrorism

Profiling and other techniques of data mining are increasingly used in the fight against transnational terrorism. Within the project DETECTER (Moeckly, 2009, D8.1), a survey was carried out to analyze and present different examples of counter-terrorism data mining systems. The survey explores some of the main projects, mainly developed in the United States (since the most prominent discussion of data mining plans and activities for counter-terrorism purposes has been in the US), where data mining techniques were used in the detection of potential terrorist and possible prevention of terrorist attacks[14].

The way in which these techniques are applied is not always straightforward. "Data mining and other forms of data analysis that are being carried out or explored in the counter-terrorism context represent one stage in a series of data-related practices, each of which presents particular issues with respect to privacy, ethics, and human rights" (Moeckli, 2008, p.2).

As far as the European Union is concerned, on 18 November 2002, Article 36 Committee of the European Union submitted a draft Council Decision which would establish terrorist profiles to be used in European counter-terrorism efforts. The document foresaw that the Member States would exchange information and cooperate both amongst themselves and with Europol to develop profiles.

The "Terrorist Rasterfahndung" is an example of anti-terrorism system put in place by the Federal Criminal Police Office in Germany. Rasterfahndung was regularly used in the 1970s as a means to tackle terrorism from the Red Army Faction - RAF, and is currently used against the Islamic terrorists (COT, 2008, p.17). Following the attacks of September, 11 the Police Office put in place a nation-wide implementation of the system, in order "to turn up the names of males between the ages of 18 and 40 who were from certain Islamic states and were either current or former students. The aim was to uncover "sleepers" who were somehow involved in terrorist activity or planning […] This use of a Rasterfahndung became the subject of a controversy before the German Constitutional Court" (Moeckli, 2008, p.35).

Data mining techniques actually allow the agencies involved in intelligence and law enforcement to work in a faster and more efficient way. "The ability of data mining to reveal associations that analysts might not think to inquire after may have also offered some hope that data mining would not only assist in performing traditional investigation tasks but could uncover connections or leads that traditional techniques would not" (Moeckli, 2008, p.10). Some of the objectives that could be covered in the counter-terrorism context are: the discovery of terrorists and terrorist networks; the generation of profiles (mainly in the flight screening context); the assessment of risks; the provision of analytic assistance. This typology is not intended to be comprehensive though, and additional type of applications in the counter-terrorism context might be of course developed in the next future (Moeckli, 2008).

[14] For an overview of data mining projects on anti-terrorism see: http://www.detecter.eu/index.php?option=com_content&view=section&layout=blog&id=7&Itemid=9

Financial domain

The attention on how the funding of terrorist activities can be detected and prevented also through the analysis of the financial system, i.e. through anti-money laundering profiling, make a connection with the following application domain of profiling techniques: the financial domain.

The application of profiling practices in the financial sector mainly concern the fight against money laundering and the prevention of fraud, as well as the broader issue of taxation.

All the different sectors such as banking, finance, accountability and the legal sectors are requested to establish procedures to facilitate the reporting of suspicious activities to the law enforcement agencies. One of the most common procedures in United States financial regulation is the so called Suspicious Activity Report (SAR). When an institution has the suspect that a customer is processing financial transfers from criminal proceeds, it is requested to prepare a SAR and to channel it to the relevant governmental agency in charge of fighting and preventing money laundering. "The use of automated monitoring systems is often seen as a powerful ally in detecting suspicious activity, justified by the wholesale increase in size of the typical transactional database […] This systems usually consist of powerful algorithms" (Canhoto, 2005, p.57). The main problems connected are, on one hand, the heavy investments in technology while, on the other, the need to employ several people to eliminate the so-called false positives.

Another related problem of the use of automated profiling in Anti-Money Laundering, as highlighted in the Fidis Project[15] research (Canhoto, 2005), is that profiles usually relies on tried and tested money laundering typologies. They do not keep up with the complex and advanced mechanisms of money laundering. Moreover, the organizations involved tend to focus their attention on the "usual suspects and give more attention to anomalous activity coming from individuals with a given demographic profile" (Canhoto, 2005, p.58).

Another important use of profiling occurs within the prevention of financial fraud. An example in this field is the German system SCHUFA, funded by national banks and other financial service providers. With the consent of their clients, the providers of bank and financial services transfer the data concerning the bank accounts and the financial behaviours to the SCHUFA. The behaviour of so-called reference-groups is then analyzed with massive data volumes. The profiling gives a scoring value, which should express the risk based on personal behaviours. This data, together with other information, are used "to determine the risk of defaulting on credit and conditions under which someone can obtain credit" (Canhoto, 2005, p.59).

Another aim of the use of profiling techniques in the financial domain is the fight against tax-evasion. The information about citizens' financial situations and their spending habits can be analyzed in order to find discrepancies and non-standards transactions. An example is the tool named "Redditometro", used by the Italian government and its "Agenzia delle Entrate" (equivalent of the British "Inland Revenue") to fight the phenomenon of tax evasion within the country. The idea behind the creation of this tool is to collect, pre-emptively,

[15] FDIS (Future of Identity in the Information Society) is a Network of Excellence supported by the European Union. See: http://www.fidis.net/

all data concerning the taxpayers and place it in a unique database. Then, specific data-mining software try to detect the "non-standard" transactions according to the parameters previously identified.

Health care domain

"In healthcare, data mining is becoming increasingly popular, if not increasingly essential" (Koh, Tan, 2005, p.64). Here, data mining techniques provide the methodology and technology to transform the huge amount of data generated by healthcare transactions into useful information for decision-making. Some application areas can be the evaluation of treatment effectiveness, management of healthcare or customer relationship management and the detection of fraud and abuse.

For example, data mining applications – as part of the profiling technical process – can be developed to evaluate the effectiveness of particular medical treatments. The comparison of the outcomes of treatments of the same disease with different medicine regimens can provide inputs on the best and most cost-effective ones. "Other data mining applications […] include associating the various side-effects of treatment, collating common symptoms to aid diagnosis, determining the most effective drug compounds for treating sub-populations that respond differently from the mainstream population to certain drugs, and determining proactive steps that can reduce the risk of affliction" (Koh, Tan, 2005, p. 64).

Of course there are also limitations. One of them is the accessibility of data, but problems may also arise if data are missing, non-standardized or corrupted. The successful application of data mining is linked both to the methodology and tools used and it requires also a good knowledge of the domain field.

Another interesting application field is the predictive medicine, which mainly deals with learning models to predict patients' health or the likelihood of a treatment being successful for a particular patient based on certain group characteristics. Usually profiling and other data mining methods are applied in clinical contexts to analyze data in order to provide healthcare professionals with the opportunity to use large amounts of data collected during their daily activities.

"Moreover, clinicians can nowadays take advantage of data mining techniques to deal with the large amount of research results obtained by molecular medicine, such as genetic or genomic signatures, which may allow transition from population based to personalized medicine" (Bellazzi, Ferrazzi, Sacchi, 2011, p.416).

In an article published on Forbes online, the director of the Stanford Genome Technology Center - Ronald W. Davis - starts with the following question: "Is genomics and personalized medicine the greatest business and investment opportunity since the advent of the Internet?" (Davis, 2012).

Of course there can be different answers and opinions on this particular issue, but what is relevant for the purposes of this essay is that there exist complex data mining and computer modelling systems, which are aimed at mapping individuals' risk factors and responses to treatments. They offer "personalized set of behaviour and treatment recommendations that can help … reduce those risks… and even prevent diseases itself" (Davis, 2012).

Employment

Within the employment sector, both at private and public level, profiling has become increasingly important for security reasons. Some areas of application are, for instance, the prevention of fraud (i.e. in the retail sector, to determine unusual cash flow and possible embezzlements); the supervision of the employees, both direct and indirect (i.e. in the postal service or the call centres to monitor work-hours); profiling on log-files and intrusion detection/prevention systems.

For instance, in Germany and Switzerland there are examples of the use of profiling techniques to detect possible embezzlement of cashiers in the supermarkets. The cash refund transactions are analyzed: whether a higher rate of refund transactions than average is detected, this could mean there is fraud. Through profiling, a target of workers is identified in order to carry out further investigations and detect possible abuses.

Another area of interest is the management of Human Resources. Data mining techniques are used in order to classify/analyze the potential and the capacities of the employees with the aim of optimizing the distribution of employees within a company. Due to the implementation of advanced profiling techniques, the workers should be informed about the functioning of the program, and they have the right to be updated on the results of their data processing. In order to better protect the rights of the employees, the use of these systems should be restricted according to different criteria, and the workers should be entitled to object or challenge the results of the scoring process in a transparent way.

Another field of security control based on profiling practices is the surveillance of Internet access and e-mail communication made by the employers both in the public and private sector. While for Human Resource management the distributive group profiling is used, in this case the profiling is personalized: intrusion detection/response systems are implemented. Besides the reporting of incidents, this application can be justified as a preventive tool in view of possible information thefts or other unlawful content-related activities from and inside the organization.

Marketing

Besides the application of dataveillance and profiling in the financial and health care domains, they are also a routine commercial practice for companies: a way of processing customers' information to devise, for instance, targeted advertisement (Jeandesboz, Bigo, Frost, 2011).

Based on the analysis of consumption patterns "in order to predict future behaviors and develop more targeted advertisement activities through the extraction of data from large sets of information" (Ivi, p.14), the commercial practices of dataveillance seek to determine the preferences of the consumers.

The reason why private companies are interested in knowing the consumers' (future) behaviours is quite obvious: improve their profits and, in general, encourage or reward those actions which are more profitable for the organization itself. To this aim, what is more interesting is not only the knowledge of the individuals' behaviours, but "to generalize from observed behaviour in order to make predictions about the behaviour of specific types of consumers" (Canhoto, 2005, p.55).

Data mining classifies users into a limited number of clusters. Based on the patterns revealed during the data mining process, profiling tries to predict / pre-empt the future behavior of the consumers, classifying the individuals in commercial "opportunities". Practical applications of these techniques in the marketing field are comprised in the "Customer Loyalty Programmes", which are structured and long-term marketing efforts aimed at providing incentives to customers who demonstrate loyal buying behaviour for example through granting them a certain amount of discount. According to Kamp, Körffer, Meints (2007), in their analysis of the German market, in addition to the data needed for discount purposes, in many cases even personal data are registered, such as the birth date, contact details, personal life related information etc. These data will be used both for market research and for advertising purposes through the use of profiling techniques.

The worldwide diffusion of the Internet and the increasing use of online social media have also permitted different forms of profiling application to the marketing domain, mainly based on web tracking behaviours and the so-called social media marketing, as we will see in the next section.

Social media and web

Paragraph 1.2 explored the issue of online behavioral profiling, also known as "targeting". As underlined by Castelluccia (2012, p. 21), "profiles are very valuable for many companies in customizing their services to suit their customers, in order to increase revenues". Through behavioral targeting online they can track the users and build profiles based on their main characteristics, interests and shopping activities. For example, behavioral profiling is used by e-commerce platforms to recommend certain products which are likely to be of interest to the users according to their particular profiles. In general, the advertising companies use this technique to display advertisements which reflect the users' interests.

Another emerging phenomenon is the so-called "ubiquitous advertising" (Castelluccia, 2012, p. 23), which means that advertising is linked both to online and physical profiles, thanks to the use of the smart phones. Since the mobile phones are usually linked to a specific person, more information can be collected and more detailed profiles can be consequently derived.

Another interesting aspect to be analyzed in the field of online profiling is the use of social media and social networks, which gained an increasing popularity in the recent years. Their characteristic is that users can make contacts and share personal information on a large scale. In order to be easily identified, people need to share information, but at the same time they have not a clear idea of who access their information and to what extent.

Social networks also play an increasing role in helping the tracking companies – which already track you over multiple websites, following you as you browse the web – to collect personal information on the users. According to Krishnamurthy, Wills (2009) "social networks such as Facebook, LinkedIn and My Space are giving the hungry cloud of tracking companies an easy way to add your name, lists of friends, and other profile information to the records they already keep on you" (Eckersley, 2009, Part 2). When the user log into a social networking site, it includes both tracking code and advertising in such a way that the third party can go to your profile page, record the contents and add the information to their files.

Besides the personal information shared by the user, a significant amount of information can also be taken from the structure of their network and group information. For instance "a study lead by MIT students, called the Gaydar project, has shown that it is possible to predict with a fairly high accuracy the sexual preferences of an individual. This is possible even if his profile is private, just by looking at the amount of gay friends it includes, compared with a person sampled randomly from the population (Castelluccia, 2012, p.8).

The diffusion of Mobile Online Social Networks (MOSN) also contributes to better track users' habitudes and information. In particular "new" MOSN, differently from traditional ones which are just adapted to the mobile context, are specifically created for the mobile format and base their content on the possibility of crossing information belonging from the location of the user's device and the position of his friends. As highlighted by Krishnamurthy and Willis, (2009) the predominant concepts of MOSN are presence and location. These information and also private data, as gender, name social networking identifier, ecc, could leak to users within the same MOSN, to users within other OSN, and even to third-party tracking sites. "Thanks to the location information, unique identifiers of devices, and traditional leakage of information now give third-party aggregation sites the capacity to build a comprehensive and dynamic portrait of mobile online social network users" (Castelluccia, 2012, p.26).

4. Profiling definition

The analysis carried out in the previous pages allows us to formulate a working definition to be used in the PROFILING project.

As underlined at the beginning of the paper, profiling here is taken into consideration as machine profiling, i.e. a process where automated decisions taken by or facilitated by machines are the core elements. But what does this mean in practical terms?

Within the academic realm, it is worth underlining that there are not many definitions. Technical literature on data mining does not analyse in depth the definition of profiling. It focuses the attention in data mining techniques and predictive models. It is mainly the socio-legal literature that provides the richest panorama of definitions.

Gary T. Marx (Marx and Reichman, 1984, p. 429) gave one of the oldest definitions of profiling in a paper that analyses systems of data searching. Profiling (defined by the author in contrast with “matching”) is defined by stressing the logic behind it: “the logic of profiling is more indirect than that of matching. It follows an inductive logic in seeking clues that will increase the probability of discovering infractions relative to random searches. Profiling permits investigators to correlate a number of distinct data items in order to assess how close a person or event comes to a predetermined characterization or model of infraction”. According to the author’s background, this definition is strictly related to the law enforcement domain.

Almost ten years later, Roger Clarke (1993, p. 2) clarifies that “profiling refers to the process of creating and using such a profile”. It can be defined as a “dataveillance technique (…) whereby a set of characteristics of a particular class of person is inferred from past experience, and data-holdings are then searched for individuals with a close fit to that set of characteristics”.

A legal scholar, Bygrave (2002, p. 301) again stressed that ‘profiling is the inference of a set of characteristics (profile) about an individual person or collective entity and the subsequent treatment of that person/entity or other persons/entities in the light of these characteristics.’

Later on, Mireille Hildebrandt was the one who put the best effort to precisely define profiling, also taking into account the technological evolution.

In “Profiling the European Citizen”, the first comprehensive book on profiling, Hildebrandt (2008a, p.41) gave the following definition: “the process of ‘discovering’ correlations between data in databases that can be used to identify and represent a human or nonhuman subject (individual or group), and/or the application of profiles (sets of correlated data) to individuate and represent a subject or to identify a subject as a member of a group or category”. She also adds that to clarify the meaning, it is worth adding the purpose of profiling: “besides individuation, profiling mainly aims for risk-assessment and/or assessment of opportunities of individual subjects”.

In another definition (Hildebrandt 2009b), the process of 'discovering' correlations between data in databases became, more precisely, "the process of 'discovering' patterns in data in databases". She also underlines the importance of prediction. Profiling is the "discovery of patterns that present knowledge which enables anticipation of future events based on what happened in the past" (Hildebrandt, 2009b, p.289).

Besides Hildebrandt, prediction is underlined also by Fuster et al. (2010, pp.1-2), who stress how profiling is commonly used "in contemporary security-related discussions as referring to the use of predictive data mining to establish recurrent patterns or 'profiles' permitting the classification of individuals into different categories". They identify a two-stage process: "a first analysis of data to look for seemingly relevant patterns, and a second examination to identify the items that correspond to the patterns".

Again Hildebrandt (2009c, p. 241) underlines that profiling "seems the only technology capable of detecting which data make a difference"; as a matter of fact "instead of mining data on the basis of predefined classes (which would produce a query that does not provide what one does not already know), profiling uses algorithms to locate unexpected correlations and patterns".

On the other hand, legal definitions stress the aspect of "automatic data processing technique" (CoE) or in other words "automated processing of personal data" (Draft report on GDPR, 2012) in order to analyse or predict habits, behaviours, preferences of an individual or a group.

Taking into account the distinctive elements analysed in section 1 (cfr. 1.1 and 1.2) and the several purposes that profiling can have, as described in section 3, the working definition used in the project, that will focus both on public and private domains of application, will be the following:

Profiling is a technique to automatically process personal and non-personal data, aimed at developing predictive knowledge from the data in the form of constructing profiles that can subsequently be applied as a basis for decision-making. A profile is a set of correlated data that represents a (human or non-human, individual or group) subject. Constructing profiles is the process of discovering unexpected patterns between data in large data sets that can be used to create profiles. Applying profiles is the process of identifying and representing a specific subject or to identify a subject as a member of a specific group or category and taking some form of decision based on this identification and representation.

References

- Amended Proposal for a Council Directive on the protection of individuals with regard to the processing of personal data and on the free movement of such data (15 October 1992), COM(92) 422 final – SYN 287.
- Article 29 Working Party (2012), Opinion 01/2012 on the data protection reform proposals, adopted on 23 March 2012.
- Article 29 Working Party (2013), Advice paper on essential elements of a definition and a provision on profiling within the EU General Data Protection Regulation, adopted on 13 May 2013.
- Bellazzi R., Ferrazzi F., Sacchi S. (September-October 2011), Predictive data mining in clinical medicine: a focus on selected methods and applications, Data Mining and Knowledge Discovery, Wiley Interdisciplinary Reviews, 1(5), pp. 416–430.
- Benetton A. (11 March 2013), Redditometro, il ricorso va a segno, The Fielder. Available at: http://thefielder.net/11/03/2013/redditometro-il-ricorso-va-a-segno/#.UWKjJlKH3iS
- Bygrave L. A. (2001), Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling, published in Computer Law3 & Security Report, "001, volume 17, pp. 17-24.
- Bygrave L. A. (2002), Data protection law: approaching its rationale, logic and limits, The Hague, Kluwer Law International.
- Brandon B. (2012), Defining 'big data' depends on who's doing the defining. When does data become big? AWS, IBM and research firms each have their own definitions. Available at: http://www.networkworld.com/news/2012/051012-big-data-259147.html
- Calders T., Custer B. (2013), What Is Data Mining and How Does It Work?, in Custers B., Calders T., Schermer B., Zarsky T. (ed.), Discrimination and Privacy in the Information Society. Data Mining and Profiling in Large Databases, Springer, pp. 27-42.
- Canhoto A. (2005), Anti-money laundering profiling, FIDIS project D. 7.2: Descriptive analysis and inventory of profiling practices, pp. 57-58.
- Castelluccia C. (2012), Behavioural Tracking on the Internet: A Technical Perspective, in Gutwirth S., Leenes R., De Hert, P., Poullet, Y., (Eds.) European Data Protection: In Good Health?, pp. 21-34.
- Cate F. H. (2008), Data Mining: The Need for a Legal Framework, Harvard Civil Rights-Civil Liberties, 43, no.2, pp. 435–489.
- Cavoukian A., Jonas J. (2012), Privacy by Design in the Age of Big Data. Available at: http://www.ipc.on.ca/english/Resources/Discussion-Papers/Discussion-Papers-Summary/?id=1195
- Clarke R.A. (1988) 'Information Technology and Dataveillance' Comm. ACM 31,5 (May 1988) Re-published in C. Dunlop and R. Kling (Eds.), 'Controversies in Computing', Academic Press, 1991.
- Clarke R. (1993), Profiling: A Hidden Challenge to the Regulation of Data Surveillance. Published in the Journal of Law and Information Science 4(2) (December 1993). Available at: https://digitalcollections.anu.edu.au/bitstream/1885/46248/31/07Paper06.pdf
- Clarke R. (1994), The Digital Persona and its Application to Data Surveillance. Published in The Information Society, 10(2) (June 1994), pp. 77-92. Available at: http://www.rogerclarke.com/DV/DigPersona.html

- Consultative Committee of the Convention 108 (T-PD) (2005), Progress report on the application of the principles of Convention 108 to the collection and processing of biometric data. Available at: http://www.coe.int/t/dghl/standardsetting/dataprotection/Reports/Biometrics_2005_en.pdf
- COT (2008) Transnational Terrorism: Theoretical approaches and policy discourse, project “Citizens and governance in a knowledge-based society”, WP 2, deliverable 3.
- Council of Europe, Recommendation CM/Rec(2010)13 of the Committee of Ministers to member states on the protection of individuals with regard to automatic processing of personal data in the context of profiling. Adopted by the Committee of Ministers on 23 November 2010 at the 1099th meeting of the Ministers’ Deputies.
- Davis R. W. (January 2012), It’s time to bet on genomics, articles on Forbes. Available at: http://www.forbes.com/sites/forbesleadershipforum/2012/06/01/its-time-to-bet-on-genomics/
- De Hert P., Bellanova R. (March 2011), Transatlantic Cooperation on Travellers’ Data Processing: From Sorting Countries to Sorting Individuals, Migration Policy Institute. Available at: http://www.migrationpolicy.org/pubs/dataprocessing-2011.pdf
- Eckersley P., 2009, How Online Tracking Companies Know Most of What You Do Online (and What Social Networks Are Doing to Help Them), part 2. Available at: https://www.eff.org/deeplinks/2009/09/online-trackers-and-social-networks
- European Commission (24 October 1995), Directive of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OJ L 281 , 23/11/1995 P. 0031 – 0050, Brussels.
- European Commission (20 July 2010), Communication from the Commission to the European Parliament and the Council: Overview of information management in the area of freedom, security and justice, COM(2010) 385 final, Brussels.
- European Commission against Racism and Intollerance (ECRI) (13 December 2002), General policy recommendation No 7 on national legislation to combat racism and racial discrimination, Council of Europe, Strasbourg.
- European Commission against Racism and Intollerance (ECRI) (29 June 2007), General policy recommendation No 11 on combating racism and racial discrimination in policing, Council of Europe, Strasbourg.
- European Data Protection Superviosr (EDPS) (25 January 2012), EDPS welcomes a “huge”step forward for data protection in Europe, but regrets inadequate rules for the police and justice area, EDPS Press release, Brussels.
- European Parliament (17 December 2012), DRAFT REPORT on the Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), COM(2012)0011 - 2012/0011 (COD), Rapporteur Albrecht J. P.
- European Union Agency for Fundametnal Rights (FRA) (2010), Towards More Effective Policing Understanding and Preventing Discriminatory Ethnic Profiling: A Guide, Luxembourg: Publications Office of the European Union.
- Europa Press Release RAPID (2 February 2011), EU Passenger Name Record (PNR)-Frequently Asked Question. Available at: http://europa.eu/rapid/press-release_MEMO-11-60_en.htm
- Fayyad U., Piatetsky-Shapiro G., Smyth P. (1996) From Data Mining to Knowledge Discovery: an Overview, In Fayyad U, Piatetsky-Shapiro G, Smyth P, Uthurusamy R. (eds) Advances in Knowledge Discovery and Data Mining. AAAI Press / MIT Press, Cambridge.
- Friedland, G., Sommer R. (2010). Cybercasing the joint: On the privacy implication of geo-tagging. Available at: http://www.icsi.berkeley.edu/pubs/networking/cybercasinghotsec10.pdf
- Fritsch L. (2008), Profiling and Location-Based Services (LBS), in Hildebrandt M., Gutwirth S.

(Eds.), Profiling the European Citizens. Cross-Disciplinary Perspectives, Springer, pp. 147-168.

[Fuster G., Gutwirth S., Erika E. (June 2010), Profiling in the European Union: A high-risk practice. INEX Policy Brief, no. 10.

[Ganeva T. (8 January 2012), 5 Things You Should Know About the FBI's Massive New Biometric Database, article on Alternet. Available at:

[http://www.alternet.org/story/153664/5_things_you_should_know_about_the_fbi%27s_massive_new_biometric_database

[Gantz J., Reinsel. D. (2011), Extracting value from chaos, IDC Iview. Available at: http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf

[Giannotti F., Pedreschi D. (2008), Mobility, Data Mining and Privacy. Geographic Knowledge Discovery, Springer.

[Hastie T., Tibshirani R., Friedman J. (2009), The Elements of Statistical Learning. Data Mining, Inference, and Prediction, Springer.

[Harcourt, B. (2007), Against prediction: Profiling, policing, and punishing in an actuarial age, University of Chicago Press, Chicago.

[Hayes B. (2010), Statewatch analysis: Spying in a see through world: the "Open Source" intelligence industry. Available at: http://database.statewatch.org/article.asp?aid=30317

[Hildebrandt M., Backhouse J. (2005), Descriptive analysis and inventory of profiling practices. In FIDIS Project Deliverable 7.2. Available at: http://www.fidis.net

[Hildebrandt M. (2006), Profiling: from Data to Knowledge. The challenges of a crucial technology, in DuD Datenschutz und Datensicherheit, 30(9), pp. 548-552.

[Hildebrandt M. (2008a), Defining profiling: new type of knowledge?, in Hildebrandt, M. Gutwirth S. (Eds.), Profiling the European Citizens, Cross-Disciplinary Perspectives, Springer, pp. 17-47.

[Hildebrandt M. (2008b), Profiling and the rule of law, in Identity in the Information Society, vol.1, no.1, pp. 55-70.

[Hildebrandt M. (2009a), Technology and the End of Law, in Claes E., Devroe W., Keirsbilck B. (Eds.), Facing the Limits of the Law, pp. 443–465.

[Hildebrandt M. (2009b), Profiling and AmI, in Rannenberg K., Royer D., Deuker A., Heidelberg (Eds), The Future of Identity in the Information Society. Challenges and Opportunities, Springer, pp. 273-310.

[Hildebrandt M. (2009c), Who is Profiling Who? Invisible Visibility, in Gutwirth S., Poullet, Y., De Hert, P., De Terwangne C., Nouwt S. (Eds), Reinventing Data Protection?, Dordrecht, Springer, pp. 239-252.

[International Civil Liberties Monitoring Group (2010), Report of the Information Clearinghouse on Border Controls and Infringements to Travellers' Rights. Available at: http://iclmg.ca/en/section/7

[Jaquet-Chiffelle D.O. (2007), Direct and Indirect Profiling in the Light of Virtual Persons, in Hildebrandt M., Gutwirth S., (Eds.), Profiling the European Citizens, Cross-Discinplinary Perspectives. Springer, pp. 55-66.

[Jeandesboz J., Bigo D., Frost M. (2011), Elements for an analysis of the history and contemporary trends of surveillance in Europe, SAPIENT project, D1.1 Smart Surveillance, State of the Art.

[Jonas J. (April 18, 2012), Big Data Q&A for the Data Protection Law and Policy Newsletter. Available at: http://jeffjonas.typepad.com/jeff_jonas/2012/04/big-data-qa-for-the-data-protection-law-and-policy-newsletter.html

[Kamp M., Körffer B., Meints M., "Profiling of customers and consumers – customer loyalty programmes and scoring practices", in Profiling the European Citizens, Springer, 2007, pp. 219-235.

[Koh H. C., and Tan G. (2005), Data Mining Applications in Healthcare, Journal of Healthcare Information Management, 19(2).

[Kranzberg M. (1986), Technology and History: “Kranzberg’s Laws”, Technology and Culture, 27(3).

[Krishnamurthy B., Wills C.E. (2009), Privacy diffusion on the web: a longitudinal perspective, In WWW ’09: Proceedings of the 18th international conference on World wide web, pp. 541-550.

[Krishnamurthy B., Wills C.E. (2009a), On the Leakage of Personally Identifiable Information via Online Social Netorks, WOSN’09, August 17, Barcelona, Spain.

[Kuner C., Cate F.H., Millard C., Svantesson D.J.B. (2012), The challenge of ‘big data’ for data protection, Editorial, International Data Privacy Law (2012) 2 (2): 47-49. Available at: http://idpl.oxfordjournals.org/content/2/2/47.extract#

[Liebenau, J., Backhouse J. (1990), Understanding Information: An Introduction, Macmillan, London.

[Marx G., Reichman N. (1984), Routinizing the Discovery of Secrets: Computers as Informants, in American Behavioral Scientist, Vol. 27, no. 4, pp.423-452.

[McAfee A., Brynjolfsson E. (2012), Big Data: The Management Revolution, Harvard Business Review. Available at: http://hbr.org/2012/10/big-data-the-management-revolution/ar/1

[Moeckli T. (2008), Counter-terrorism data mining: legal analysis and best practices, DETECTER project, WP6, deliverable 8(3).

[Möller J. and Florax B.-J. (2002), “Kreditwirtschafliche Scoringverfahren”, in Multimedia und recht, (12), 806-811.

[Novek E., Sinha N., and Gandy O. (1990) The value of your name, in Media Culture Society vol. 12, pp. 525-543.

[Out-Law (2013), “Albrecht’s proposed health data research processing restrictions near-identical to EDRi lobby papers”. Available at: http://www.out-law.com/articles/2013/february/albrechts-proposed-health-data-research-processing-restrictions-near-identical-to-edri-lobby-papers/

[Papakostantinou V., De Hert P. (June 2009), The PNR Agreement and Transatlantic anti-terrorism Cooperation: No firm human rights framework on either side of the Atlantic, in Common Market Law Review, vol. 46 n.3.

[Proposal for a Council Directive concerning the protection of individuals in relation to the processing of personal data, (COM (90) 314 final – SYN 287, 13.9.1990.

[Proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), SEC(2012) 72 final, Brussels, 25.1.2012 COM(2012) 11 final 2012/0011 (COD).

[Roosendaal A. (2010), Digital Personae and profiles as Representations of Individuals, published in Privacy and Identity management for life, p.226–236

[Roosendaal A. (2012), We Are All Connected to Facebook . . . by Facebook!, in Gutwirth S., Leenes R., De Hert, P., Poullet Y. (Eds.), European Data Protection: In Good Health?, pp. 3-20.

[Roosendaal A. (2013), Digital Personae and Profiles in Law. Protecting Individuals’ Rights in Online Contexts, Oisterwijk: Wolf Legal Publishers.

[Siegele L. (2 December 2011), Welcome to the yotta world, The Outlook for 2012, Economist. Available at: http://www.economist.com/node/21537922

[United Nations General Assembly (2007), Report of the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism, Martin Scheinin (A/HRC/4/26), §§ 34 and 41.

[US military handbook (5 December 2006), Open Source Intelligence. Available at: https://

www.fas.org/irp/doddir/army/fmi2-22-9.pdf

[USI, Trilateral, CSSC (2011), Emerging Smart Surveillance Technologies, in SAPIENT project, D1.1, Smart Surveillance: State of the Art.

[Valigra L. (February 2012), Personal genomes hold eventual promise for treatments, Boston Business Journal, 9. Available at: http://www.bizjournals.com/boston/blog/mass-high-tech/2012/02/personal-genomes-hold-eventual-promise.html?page=all

[Van Brakel R., De Hert P. (2011-2013), Policing, surveillance and law in a pre-crime society: understanding the consequences of technology based strategy, CPS, n.20.

[Vedder A.H. (1999) KDD: The challenge to individualism, in Ethics and Information Technology, no. 1, pp. 275-281.

[Yannopoulos A, Andronikou V, Varvarigou T. (2008), Behavioural Biometric Profiling and Ambient Intelligence, in Hildebrandt M, Gutwirth S (eds.), Profiling the European Citizen. Cross-Disciplinary Perspectives, Springer, pp. 109-131.

[Zarsky T. Z. (2002-2003), `Mine Your Own Business!`: Making The Case For The Implications Of The Data Mining Of Personal Information In The Forum Of Public Opinion.” Yale Journal of Law & Technology 5, pp. 1-56.

[Zarsky T. Z. (2011), Governmental Data Mining and its Alternatives, Penn State Law Review, 11(2) pp. 285-330.

prof[L]ng

PROTECTING CITIZENS' RIGHTS FIGHTING ILLICIT PROFILING