# Digital Personae and Profiles as Representations of Individuals

Arnold Roosendaal



# Digital Personae and Profiles as Representations of Individuals

Arnold Roosendaal

Tilburg Institute for Law, Technology, and Society (TILT)
Tilburg University
PO Box 90153, 5000 LE Tilburg, The Netherlands
A.P.C.Roosendaal@uvt.nl

**Abstract.** This paper explores the concepts of digital personae and profiles and the way they represent individuals. Even though their manifestation as data sets seems similar, they originate in different ways. The differences between the two forms of digital representations have major implications for their connection and application to known individuals. Digital personae are connected to known individuals in the real world, whereas profiles are not. However, different types of identification can establish the connection between a profile and an offline individual. A profile can then transform into a digital persona. The differences between digital personae and profiles have implications for the applicability of data protection regulations and influence the amount of control individuals have over their representations and decisions based on these. This paper shows the relation between digital personae and profiles and indicates where privacy and autonomy of individuals can be at stake.

**Keywords:** Digital Persona, Profile, Representation, Individual, Data sets

## 1 Introduction

The enormous amount of electronic data inherent to the information society facilitates the establishment of digital personae [1], representations of individuals in the form of data sets. These digital personae are used by governments or businesses to take decisions that affect the represented individual. Digital personae are consciously created with a specific, indicated purpose, and the concerned individual is usually aware of the representation being created. Another form of digital representations are profiles. These are the result of automated processes where large data sets are processed in order to arrive at (a set of) characteristics which can be used as a basis for decision making. Usually, in particular in the case of group profiles, the represented individual is not known in the real world beforehand, but a profile can be connected to a known individual later on.

This paper presents the concept of a digital persona (section 2) and of a profile (section 3) and explores similarities and differences between the two (section 4). It appears that the manifestation of both forms is basically similar, namely as a data set comprising attributes instantiated with values associated to the individual, but the

differences in the way they are constructed and the intended purpose and connection to individuals in the real world are essential to gain further insight in how the represented individuals are affected. Section 5 analyses this connection between individuals and data sets from a legal perspective. The real world individuals are the underlying entities which are represented by data sets (identities) [1]. These data sets can contain personal data. Personal data means: "any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity" (Art. 2(a) Data Protection Directive (DPD)[1]). Thus, for the applicability of the DPD it is important to know whether the connection between a digital persona or a profile and the underlying entity can be made based on the data in the representation. It appears that the DPD might be applicable in the case of a digital persona as well as in the case of a profile as a basis for taking decisions. The main focus is to clarify to what extent digital personae and profiles can be connected to entities and to come to a common understanding of the two concepts. In section 6 the conclusions are drawn.

## 2 Digital Personae

A digital persona is a representation of an individual, identifiable[2] by the one who creates and/or uses the data set. The concept of a digital persona was introduced by Roger Clarke, who used the following definition: "a model of an individual's public personality based on data and maintained by transactions, and intended for use as a proxy for the individual" [2]. The representational capacity is a key element. It follows from the definition that functioning as a proxy for a specific individual is intended, so the representations that qualify as a digital persona are limited to those data sets which contain an identifying link to an entity. To compare, Solove, for instance, takes a much broader perspective when he talks about a digital person. He states that "it is ever more possible to create an electronic collage that covers much of a person's life – a life captured in records, a digital person composed in the collective computer networks of the world" [3]. Solove's digital person includes digital personae as well as profiles, which will be discussed later on in this paper, and other data sets. In the case of a digital persona, the purpose of its creation is known beforehand, and therefore the data that are needed to form the representation are also known or at least to a certain extent. This implies that creating a digital persona can be compared to filling out a template since it is known which attributes one needs.

Clarke distinguishes between projected personae and imposed personae. A projected digital persona is "an image of one's self that an individual conveys to others by means of data", for instance by creating a personal page on a social network site, whereas the imposed digital persona is "an identity projected onto a person by

[1] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, No L 281/31.

[2] Identifiability can take different forms. See below, section 3.1.

means of data, by outside agencies such as corporations and government agencies" [2], for instance a record created by a credit rating agency. A combined form is also possible, for instance when an electronic patient record (usually called a 'profile') is created. The concerned individual is closely involved in the creation and provides a major part of the data. The health provider stores the data and adds personal interpretations and other data (e.g. diagnoses and personal observations). The creation and maintenance of the digital persona is based on transactions, which can be any kind of interaction between the concerned individual and persons or technical devices.

The data that form a digital persona can function as or are a representation of a partial identity of the individual. A partial identity is a subset of attributes of a complete identity, where a complete identity is the union of all attributes of all identities of this person [4]. Usually, a digital persona is created for use in a specific context, so the data that are relevant for the purpose are limited to this context. For instance, data concerning the income and taxations of an individual are not relevant for a medical dossier, so they should not be included there. Even though the represented individual is aware of the existence of digital personae, she does not always know what the contents exactly are. In particular in the case of imposed personae, the individual may be aware of part of the data, mainly those data that are obvious to be included, such as name and address and specific context related data, but the individual may not know which additional data are part of the representation (e.g. a medical diagnosis).

## 3 Profiles

Another form of digital representations of individuals are profiles. These are the result of an automated process where large data sets are processed in order to come to (a set of) characteristics which can be used as a basis for decision making. A profile is a set of correlated data which is created with the use of profiling technologies, a set of technologies with as a common characteristic the use of algorithms or other techniques to create, discover or construct knowledge from huge sets of data. Profiling can be defined as "[t]he 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" [5] or the creation of a representation based on automated monitoring of individual behaviour. The data can be aggregated from different sources. In first instance, there is no direct connection to an entity, so individuals that can be affected later on are not (necessarily) aware of the data collection.

Profiles concern groups or individuals. Group profiles describe a set of attributes concerning a group of people and are created with a data mining process. Group profiles can be distributive or non-distributive. In the case of a distributive group profile, the attributes of the group are also the attributes of all the members of the group. For instance, the attribute of 'not being married' for a group of bachelors also counts for each individual member of the group. For non-distributive group profiles,

matters are more complicated. Consider again the group of bachelors, and suppose an indication is added that this group has a higher risk of getting a liver disease. This higher risk applies to the group, but not to each individual, because other factors, like drinking behaviour, are also relevant. The association is statistical rather than determinate. Here, the information contained in the profile envisages individuals as members of groups; it does not envisage the individuals as such [6].

In the case of an individual profile automatic monitoring processes are executed to collect and analyse data from a specific individual. This individual does not have to be identified (yet) when data is added to the profile, but only recognized, for instance based on a cookie. The profile is created based on monitoring behaviour of the concerned individual.

The table below gives an overview of the main characteristics of digital personae and profiles. As can be seen, the main differences between the two lie in the creation and whether the represented individual is aware of the data set. A profile can be connected to an individual later on, while the connection between a digital persona and an individual is ingrained beforehand.

**Table 1.** Characteristics of digital personae and profiles.

| **Characteristics** | **Digital Persona** | | **Profile** | |
|---|---|---|---|---|
| **Creation** | Desired attributes in 'template' | Projected persona | Result of profiling technologies: automated process | Distributive profile |
| | | Imposed persona | | Non-distributive profile |
| | | | | Individual profile |
| **Awareness** | Individual is aware | | Individual is not (necessarily) aware | |
| **Connection to individual** | Ingrained beforehand | | Can be connected/applied to a specific individual later on | |

### 3.1 From Profile to Digital Persona

Even though there is no direct connection to a specific entity, a profile can be connected to or applied to an individual later on. The connection to an individual can be made based on the identification of an individual as having one or more attributes contained in the profile. Leenes [7] distinguishes between different forms of identifiability. Depending on the data in the data set, in his terms, the identifiability can be L-identifiability for Look-up identifiability or R-identifiability for Recognition identifiability. L-identifiability means that there is a register or table that provides the connection between an identifier and an individual, such as a phone directory which links phone numbers to names. In case of a digital persona, the data set always contains an L-identifier, like a name or a passport number. This implies that there is a direct connection to an individual and that data protection regulation applies.

Profiles do not contain L-identifiers, but they connect to individuals in an indirect manner. As seen above, an individual profile may contain an R-identifier, such as a cookie, which facilitates the recognition of the individual when she returns to the site

of the profiling one (e.g. Amazon). A group profile refers to a number of people. People that show certain behavior or an attribute that is in the profile can be identified as belonging to a certain class. After recognition as a member of a group, an identifier can be issued to enable R-identification in the future. So, according to Leenes [7], the typical procedure will be: after the group profile is instantiated to the individual an R-identifier (e.g. cookie) is issued to the individual to maintain the link. The group profile is now an individual profile. It is important to note that at this point (R-ID in profile) there is no link to an entity.[3]

An individual profile can become a digital persona when an L-identifier is added. For instance, an individual at a certain point in time gives identifying information, or the information is obtained from another source. The L-identifier makes the connection between the individual profile and an offline individual. Since the data in the profile is provided by a third party it takes the form of an imposed digital persona. With regard to data protection, group profiles are excluded. Individual profiles, however, are in a grey area, because there can be discussion on whether an R-Identifier can indirectly identify an individual. An example of such a discussion can be found in IP-addresses [9]. The figure below gives a schematic overview of the relation between profiles and digital personae.

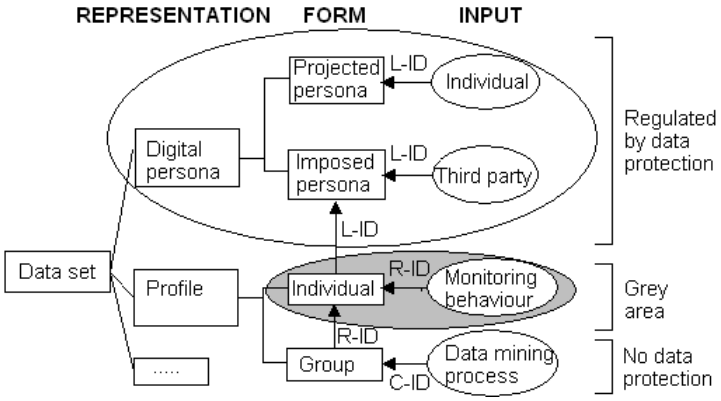


**Fig. 1.** The relation between digital personae and profiles. The C-ID is a non-individual identifier as belonging to a class and applies to all individuals in the group.

## 4 Digital Personae and Profiles: Similarities and Differences

This section describes the similarities and differences between digital personae and profiles. As shown above, their main differences lie in the way they originate and in the link to an individual. A digital persona is created with the aim of representing a specific known individual and often the concerned individual herself is involved in

[3] The used theory as developed by Leenes is helpful to distinguish between different identifiers. To calculate the probability of an R-identifier, additional tools, such as the Shannon/Weaver theory [8], are needed. This paper is, however, not on information theory, so that complementing aspect is not included here.

providing (parts of) the data in the digital persona. A profile is usually created with profiling technologies out of a set of aggregated data and is meant to reveal patterns. A profile refers to a group of people or to an individual without identification. After the identification of an individual that fits the profile the individual profile becomes an imposed digital persona.

Profiles as well as digital personae are meant as representations. Whether they are capable of representing a known individual or not distinguishes the one from the other, but they both have representational capacity. Presenting something in text or images is always a form of representation, since it refers to an original (absent) object. How this representation works can be explained with the help of semiotics, in particular the theory of the 'triad of meaning' as developed by C.S. Peirce. His triad is a model of how things get meaning [10]. There are conflicting views on this triadic theory, including proposed adaptations to the model. For instance, there have been proposals for a category of Fourthness which question the sufficiency of Peirce's semiotic, and proposals for a reduction to dyadicity which would render the semiotic triad unnecessary.[4] However, the aim of this paper is not to set out semiotic theory and the different possible viewpoints. Since Peirce's triadic model is widely accepted, I take this model as a starting point for illustrating my view on representation and the differences between digital personae and profiles. According to Peirce, the process of ascribing meaning to a certain object is always an interactive process between three things: the object, the sign, and the interpretant. The object is the thing to which a certain meaning, the knowledge of the object at a specific moment (the interpretant[5]), is ascribed. This object can be anything, physical as well as virtual. The only precondition is that the receiver of information that leads to the interpretant is able to have an idea about the object, for instance based on past experiences. The sign is something that stands for the object, since it is impossible to have knowledge on an object in a direct manner. "The sign is an instruction for interpretation, a mechanism which starts from an initial stimulus and leads to all its illative consequences" [11]. This implies that for each person the interpretant can be different, since the sign is interpreted and this interpretation can lead to different outcomes. Peirce's theory can be visualised as follows:

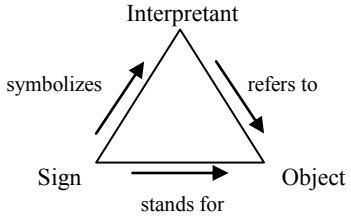


**Fig. 2.** Peirce's Triad of Meaning

When applied to the situation of a digital persona related to an individual the triad can be filled in as follows:

[4] See, for instance: http://www.paulburgess.org/triadic.html.

[5] The interpretant is an interpretation in the sense of the result of the process of interpretation. It is formed in the mind of the receiver of the information.

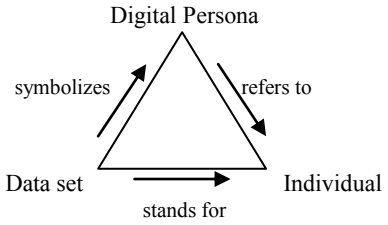


**Fig. 3.** The Triad of Meaning applied to a Digital Persona

Here, the individual is the object, the element to which a certain meaning is ascribed. The data set is the sign that there is an individual and shows information which can be interpreted and leads to the interpretant, a digital persona. The interpretant has to reveal the knowledge concerning the individual at a certain moment. The digital persona can become the starting point for a new semiotic process in the function of a new sign. This sign is interpreted and leads to a new interpretant and further knowledge about the original object, the individual.

Now, consider the same process with the digital persona replaced by a (distributive) profile. In this case, the data set can be interpreted, leading to a profile. The data, however, are now related to an unknown or potential individual instead of to a known individual, known to the one who interprets the data set, as is the case with a digital persona. Once the individual is known, the profile can become an imposed digital persona in the sense that the individual is considered to be in conformity with the profile. It is an image projected onto a person by others.

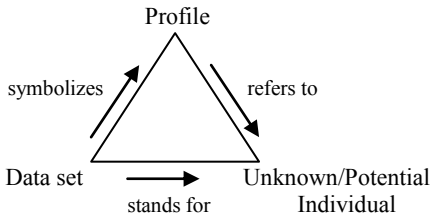


**Fig. 4.** The Triad of Meaning applied to a Profile

A digital persona stems from data that are directly related to and coming from a specific individual. A group profile stems from data that are collected from numerous individuals and forms an image that might be applicable to one or more of the individuals in the group. It appears that digital personae can be seen as explicit representations of individuals, whereas profiles are implicit, or more indirect, representations. Nevertheless, the manifestation of both is similar; a data set. The major difference lies in how meaning is ascribed to the individual. In the case of a digital persona, the meaning is ingrained beforehand, while in the case of a profile certain attributes or patterns can reveal information. Due to the differences, profiles and digital personae should be treated differently by those who use the representations

as a basis for taking decisions concerning individuals, although their manifestation as a data set is similar.

This statement will be discussed now from a legal perspective. So, the next section will explore whether the applicability of the DPD really is dependent on whether a data set is a digital persona or a profile.

## 5 Legal Embedding

Essentially, regardless whether one is dealing with a digital persona or a profile, individuals can be affected by decisions that are taken based on the data sets. When personal data are involved in the processing, data protection legislation applies. At a European level, this means that the data processing has to comply with the provisions of Directive 95/46/EC (DPD). With regard to decisions taken based on the processing of (personal) data, Article 15 of the DPD is very relevant. It grants the right to every person not to be subjected to decisions that are taken based solely on the automated processing of data. So, 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 [12].

In industry and commerce, automated decision-making is common practice [13]. This is not strange in our modern society where data and information are important assets and where automation is a standard business process. 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 on which a decision can be based. This implies that, usually, personal data are at stake in the processing. Then, the decision is based on a digital persona. However, even in the case of profiles the DPD might be applicable. 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.

The core problem is that identifiability is difficult to define. In the grey area (see section 3.1 above), where personalised profiles are at stake, but the only identifier is an R-identifier which establishes recognition as the same person, the decision will be applied to an individual. The characteristics in the profile may be too general to speak of personal data when not connected to a known individual. However, the R-identifier establishes the connection and makes that a decision, based on these (personal) characteristics can be applied to an individual. For instance, an online store recognizes a visitor and knows some general preferences. Based on earlier visits, where the person was recognized because of an issued cookie, a profile is created that shows that this person is interested in heavy metal music and books about fishing. Based on this profile, it is decided (in an automated manner) that this person receives an online offer of price reduced tickets for a heavy metal concert. In this example, the individual is affected in a positive way by the decision, but, obviously, there can often

be negative effects, for instance when someone is excluded from a price reduction, because she buys her heavy metal music at another store. Nevertheless, being affected in a positive or negative way is not the key issue. The key issue is that individuals are affected, even when their names are not known. Because the decisions are applied to individuals, perhaps even without processing personal data in a strict sense[6], the DPD should apply.

The previous paragraph had the implied assumption that there is one single user bound to a computer. This is, obviously, not completely true, since often computers are shared with a family or colleagues. However, technological development makes that electronic devices become more and more personal. Smartphones and laptops allow Internet access, regardless of one's location, and are usually used by only one individual. Besides, even when a computer is used by more than one individual, it is still possible to distinguish between the different users. Clicking behaviour and web analysis reveal patterns that relate to individuals, simply by comparing click trails and visited web sites. After a certain amount of information is revealed a fingerprint threshold is met which enables the identification (recognition) of an individual user [14].

The opposite of personalization is possible as well. Individuals can choose for so-called deliberate disinformation, which basically means that individual identifiers, such as a bar code or customer number, are posted on the Internet, allowing others to use it. When a number of individuals is using the same identifier it is no longer personal and opportunities to make appropriate individual profiles are blocked. Nevertheless, this practice can occur in the case of identifiers issued by companies, but in ordinary circumstances IP addresses and account data or login details reveal whether one is dealing with the same individual, or at least a restricted number of individuals, such as a family.

Article 15 of the DPD is meant to protect individuals from decisions being taken about them without any human involvement. This, because the lack of a human factor was deemed to be conflicting with human dignity. Another function of the DPD is to ensure transparency towards data subjects as supported by the information duties laid down in Articles 10 and 11 of the DPD. Since it was concluded that even the use of anonymous profiles as a basis for decision taking lead to affected individuals afterwards, this automated decision-making is not allowed at all, because it conflicts with the DPD. Whether the regime is meant to be so strict has to be researched further, but at least there is an important issue concerning the way data are processed in today's society. In any case, this section showed that the distinction between digital personae and profiles in the light of automated decision-making is not so relevant, even though public (and academic) debate focuses on the scope of the term 'personal data' as determining whether the DPD is applicable in a certain case or not.

Deciding that the DPD is applicable to all processing of data in the form of digital personae as well as profiles would have major consequences for the information society, which might not be the most desirable. Besides, it is always important to read and interpret legal texts while keeping an eye on the context to which the provisions are applied. This context is nowadays a different one than the context in 1995, when the DPD was written. However, research is needed to find out when the DPD should

[6] Unless the cookie is considered to be personal data, but that is a discussion on itself.

apply and when not. As long as there is no clarity, the protection goals of the DPD may not be achieved. The individual has to be the central factor around which data processing and data protection takes place. That means that the changing technologies should not be leading in deciding whether the DPD is applicable or not.

## 6 Conclusion

This paper described the concepts of a digital persona and a profile. Both are forms of representations that are used by governments and businesses to take decisions. However, there are some important differences between the two concepts, which also have implications for the way they possibly affect represented individuals. A main difference lies in the connection to a known individual and whether this connection is made before or after the representation is created. A digital persona is a direct, explicit representation, whereas a profile usually represents a group and reveals attributes that may be applicable to individuals in the group, or the profile represents an individual whose behaviour is monitored. However, the concerned individuals are not identified.

Digital personae and profiles both consist of data. Thus, their basic manifestation is similar. However, the individualisation of a data set and the way the data are collected may imply differences in the impact of the application of the representations. An important aspect is the awareness of the concerned individual of the data set being created. Without awareness, as is the case with profiles, the individual cannot influence the way the data set is used for decision taking. Another important aspect is whether individuals can exercise rights from data protection regulations. A digital persona always contains an L-identifier which establishes the connection to an offline individual, so the data in the digital persona do qualify as personal data. In group profiles this is not the case. Individual profiles are somewhat unclear in this respect, because they may very well facilitate identification, even though there is no L-identifier included.

In the end, individuals are affected by decisions taken based on the data sets. Important questions are whether it is problematic that some parts of the data processing are not regulated by data protection regulations, and whether there is a significant difference for the individual between a profile and a digital persona as a starting point of a digital representation. It is important to know how privacy and autonomy of the represented individuals are affected by these decisions and the way the representations are made. Privacy is in this context related to the applicability of data protection regulations. Autonomy relates to the amount of control an individual has in the establishment and processing of her data set and informational self-determination. This paper clarified the concepts of digital personae and profiles and their relations in order to enable further research on these implications for individuals. It also became clear that in a strict sense the DPD might be applicable to all data processing aiming at automated decision-making, regardless of whether digital personae or anonymous profiles are used as input. Applying the DPD to all processing might have major, probably undesirable, consequences for the way industry and

commerce are organized. Further research is needed in order to find out whether the DPD currently should be interpreted as including these types of data processing. A general factor in this research should be that the DPD gives certain rights to individuals to protect them. Developments in technology should not lead to the case that the DPD is not applied, while individuals and their rights are influenced anyway.

The author wants to thank prof. dr. Ronald Leenes for his stimulating supervision, as well as the reviewers for their comments on the draft version of this paper. Part of the research leading to these results has received funding from the European Community's Seventh Framework programme (FP7/2007-2013) under grant agreement No. 216483.

## References

1. Clarke, R.: Authentication Re-visited: How Public Key Infrastructure Could Yet Prosper. 16th Bled eCommerce Conference eTransformation, Bled, Slovenia (2003) 632-648
2. Clarke, R.: The Digital Persona and it's Application to Data Surveillance. The Information Society **10** (1994)
3. Solove, D.J.: The Digital Person; technology and privacy in the information age. New York University Press, New York (2004)
4. Pfitzmann, A., Hansen, M.: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management - A Consolidated Proposal for Terminology. TUD/ULD, Dresden/Kiel (2008)
5. Hildebrandt, M.: Defining Profiling: A New Type of Knowledge? In: Hildebrandt, M., Gutwirth, S. (eds.): Profiling the European Citizen; Cross-Disciplinary Perspectives. Springer (2008) 17-45
6. Vedder, A.H.: KDD, Privacy, Individuality, and Fairness. In: Spinello, R.A., Tavani, H.T. (eds.): Readings in CyberEthics. Jones and Bartlett Publishers, Sudbury Massachusetts (2004) 462-470
7. Leenes, R.: Do They Know Me? Deconstructing identifiability. University of Ottawa Law & Technology Journal **4** (2008)
8. Shannon, C.E., Weaver, W.: A Mathematical Theory of Communication. Bell System Technical Journal **27** (1948) 379-423, 623-656
9. Article 29 Working Party: Opinion 4/2007 on the Concept of Personal Data. Vol. WP136 (2007)
10. Driel, H.v.: Het semiotisch pragmatisme van Charles S. Peirce. Benjamins, Amsterdam (1991)
11. Eco, U.: Semiotics and the Philosophy of Language. Indiana University Press, Bloomington (1984)
12. Cuijpers, C.M.C.K.: Privacyrecht of privaatrecht? Een privaatrechtelijk alternatief voor de implementatie van de Europese privacyrichtlijn. Wolf Legal Publishers, Tilburg (2004)

13. Leenes, R.E.: Reply: Addressing the Obscurity of Data Clouds In: Hildebrandt, M., Gutwirth, S. (eds.): Profiling the European Citizen; Cross-Disciplinary Perspectives. Springer (2008) 293-300
14. Conti, G.: Googling Security; How much does Google know about you? Addison-Wesley, Boston (2009)