The rise of identifiability

Cuijpers, C.M.K.C.; Saygin, Y.

Published in:
APC Conference proceedings

Publication date:
2012

Document Version
Peer reviewed version

The rise of identifiability: the downfall of personal data protection?

Colette Cuijpers and Yücel Saygin[1]

Abstract

Directive 95/46/EC is applicable when ´personal data are being processed´ meaning that all data that relate to an identified or identifiable natural person, from creation to destruction, fall within the scope of the Directive. Problematic in this respect is that recent technological developments make identification on the basis of trivial information rather easy. The Article 29 Working Party has clarified identifiability depending on "the means likely reasonably to be used to identify a person" and notes that identification is possible on the basis of "the combination of scattered information". Some privacy related scandals in recent years are an example of how some advanced analysis techniques could be used to infer identities from (supposedly) anonymized data sets. In case of location data this is even more serious since plenty of background information exists which can be linked and analyzed with powerful data mining techniques. Existing research clearly demonstrates how geolocation information can fairly easily be combined with other publicly available information, turning it into identifiable and thus personal information. The Article 29 Working Party has acknowledged this characteristic of geolocation information. In this article we will voice the concern that an extensive interpretation of the concept of personal data might overshoot its purpose of enhancing data protection.

1. Introduction

Directive 95/46/EC[2] is applicable when "personal data are being processed" meaning that all data that relate to an identified or identifiable natural person - from creation to destruction - fall within the scope of the Directive. Therefore, the processing of these data must comply with all the requirements laid down in this Directive. Problematic in this respect is that recent technological developments make identification on the basis of trivial information rather easy. Moreover, the Article 29 Data Protection Working Party[3] (hereafter: Art. 29 WP) has given a very broad interpretation of personal data.

Some privacy related scandals in recent years clearly showed how some advanced analysis techniques could be used to infer identities from (supposedly) anonymized

---

[1] Colette Cuijpers is assistant professor at TILT - Tilburg Institute for Law, Technology, and Society. Yücel Saygin is associate professor with the Faculty of Engineering and Natural Sciences at Sabanci University in Istanbul, Turkey.

[2] 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. Official Journal L 281, 23/11/1995 P. 0031 – 0050. Website: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML

[3] This Working Party was set up under Article 29 of Directive 95/46/EC. It is an independent European advisory body on data protection and privacy. Website: http://ec.europa.eu/justice_home/fsj/privacy/index_en.htm

data sets. In the AOL scandal for example, only successive user search queries were released without any identifier, however individuals were shown to be identified from such data. In case of location data this is even more serious due to rich background information, which can be linked and analyzed with powerful data mining techniques. In fact in recent work it has been shown that even pairwise distances among locations could be used with triangulation and other means to find where these locations correspond to in a map, after which individuals could be identified through inferring their home or work place. The aim of this paper is to demonstrate how geolocation information can fairly easily be combined with other publicly available information, turning it into identifiable and thus personal information. The Article 29 Working Party has acknowledged this characteristic of geolocation information. In this article we will voice the concern that an extensive interpretation of the concept of personal data might overshoot its purpose of enhancing data protection. In section 2 we will elaborate upon the interpretation of the concept of personal data, especially in view of geo information, in Data Protection terminology better known as location data.[4] We will analyze the Opinions of the Art. 29 WP and have a brief look at the 2012 proposal for a Data Protection Regulation. The reason to focus on location data relates to the case we want to present in section 3. This case demonstrates the ease of identifiability, the core notion in the concept of personal data, on the basis of location data. To conclude, we will discuss in section 4 how the means - an extensive interpretation of the concept of personal data - might overshoot its purpose of enhancing data protection.

2. **Personal data**

2.1 Definition

Personal data are defined in article 2(a) of Directive 95/46/EC as:
"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."

In the preamble an interesting clarification can be found in point 26:
"Whereas the principles of protection must apply to any information concerning an identified or identifiable person; whereas, to determine whether a person is identifiable, account should be taken of ***all the means likely reasonably to be used*** *either by the controller or by* ***any other person to identify the said person****; whereas the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable; whereas codes of conduct within the meaning of Article 27 may be a useful instrument for providing guidance as to the ways in which*

[4] This concept is defined in Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (hereafter: ePrivacy Directive), Official Journal L 201, 31/07/2002 P. 0037 – 0047. Website: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32002L0058:en:HTML. It is important to note that this Directive has been amended by Directive 2009/136/EC, Official Journal L 337, 18/12/2009 P. 0011 - 0036.

data may be rendered anonymous and retained in a form in which identification of the data subject is no longer possible." (Emphasis added).

The emphasized phrases indicate that identification is not restricted to the controller or processor engaged in the processing of personal data, but relates to any means reasonably likely to be used by any person.

2.2 Interpretation of Art. 29 WP

Because in practice the scope of the concept of personal data raised all sorts of questions, the Art. 29 WP presented in June 2007 an Opinion completely dedicated to this concept.[5] It is explicitly stated that: *"it is the intention of the European lawmaker to have a wide notion of personal data"*.[6] However, there is a restriction in view of the objective, which is: "*to protect the fundamental rights and freedoms of natural persons and in particular their right to privacy, with regard to the processing of personal data*".[7] In this respect it is even acknowledged that: *"the scope of the data protection rules should not be overstretched"*.[8] Moreover it is noted that: "*it would be an undesirable result to end up applying data protection rules to situations which were not intended to be covered by those rules and for which they were not designed by the legislator"*.[9] But still in the end, the position is taken that: *"it is a better option not to unduly restrict the interpretation of the definition of personal data but rather to note that there is considerable flexibility in the application of the rules to the data"*.[10]

The Art. 29 WP subsequently clarifies the four key building blocks of the definition of personal data: *any information, relating to, an identified or indentifiable, natural person.* In relation to *any information* the main points of clarification concern the inclusion of both objective and subjective (e.g. opinions) information and the irrelevant nature of the format in which the information is kept. *Relating to* is a more difficult concept. The Art. 29 WP explains this concept by explaining three elements, of which one must be met in order for information to be *relating to.* The Art. 29 WP states in this respect: *"(...) it could be pointed out that, in order to consider that the data "relate" to an individual, a "content" element OR a "purpose" element OR a "result" element should be present"*.[11] With content it is meant that the information is about a natural person. Purpose indicates when data are used with the intent to: *"evaluate, treat in a certain way or influence the status or behaviour of an individual"*.[12] To conclude the element of result is met when the use of the data is: *"likely to have*

[5] Art. 29 WP, Opinion 4/2007 on the concept of personal data, adopted on 20th June 2007, 01248/07/EN, WP 136.
Website: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf
[6] WP 136, p. 4.
[7] Idem.
[8] WP 136, p. 5.
[9] Idem.
[10] Idem.
[11] WP 136, p. 10.
[12] Idem.

an impact on a certain person's rights and interests".[13] The concept of *natural person* is not that interesting in view of this paper. It concerns the question whether the concept of personal data also includes deceased persons, unborn children and legal persons. This is left to the discretion of the Member States. The main reason why the concept of personal data is widening relates to the fourth key building block: *identifiable*. A person is *identified* when he can be distinguished within a group of persons. In case of *identifiability* this is not yet the case, however, it might be *possible,* e.g. by linking different data sets.

Even though intended as a clarification, the general guidelines provided for by the Art. 29 WP do not really contribute to understanding the way in which the concept of personal data must be applied in practice and does definitely not prevent differences in interpretation. Broad application of the concept of personal data is still the main rule, where application of the data protection rules to the data can be flexible. This seems to contradict the whole purpose of the concept of personal data, which is to determine when the rules of the Data Protection Directive must be applied. In an attempt to connect theory to practice, the Art. 29 WP provides several real life examples. These examples demonstrate – even more clearly than the general guidelines – how extensive the interpretation of the concept of personal data in practice should be.

2.3 Uncertainty = personal data

Looking at the examples of IP addresses and camera surveillance it becomes clear that the Art. 29 WP seems to broaden the scope of *personal data* to include what we call *´uncertain data´*. This notion describes how data need to be considered to be personal data even when identifiability is uncertain. In relation to IP addresses the Art. 29 WP refers to an Internet café where users are not necessarily registered. In relation to the question whether or not in such circumstances IP Addresses are personal data, the remark is made that: *"Unless the Internet Service Provider is in a position to distinguish with absolute certainty that the data correspond to users that cannot be identified, it will have to treat all IP information as personal data, to be on the safe side".*[14] In relation to video camera surveillance a similar reasoning is presented: *"As the purpose of video surveillance is, however, to identify the persons to be seen in the video images in all cases where such identification is deemed necessary by the controller, the whole application as such has to be considered as processing data about identifiable persons, even if some persons recorded are not identifiable in practice"*.[15] In the next section we will discuss how this criteria of uncertainty is also used in relation to the personal character of location data.

2.4 Location data

Location data are defined in Art. 2(c) of Directive 2002/58/EC as: *"any data processed in an electronic communications network, indicating the geographic position*

[13] WP 136, p. 11.
[14] WP 136, p. 17.
[15] WP 136, p. 16.

of the terminal equipment of a user of a publicly available electronic communications service".

Already in several opinions, the Art. 29 WP has interpreted location data as always relating to an identified or identifiable natural person, and thus as being personal data subject to the provisions laid down in Directive 95/46/EC."[16]

The above means that to location data both regimes of Directive 95/46/EC and Directive 2002/58/EC apply. In this scenario all the general rules of the *lex generalis* (95/46/EC) apply, unless the *lex specialis* (2002/58/EC) provides for specific rules. The main difference between the two regimes is the legal ground for processing. Art. 7 of Directive 95/46/EC presents several grounds - even the legitimate interest of the data processor if not outweighed by the interest of the data subject - while the ePrivacy Directive only allows the processing of location data *"when they are made anonymous, or with the consent of the users or subscribers"*.[17]

Also in relation to location data, the Art. 29 WP has used the notion of uncertainty to qualify data as personal data: *"The fact that in some cases the owner of the device currently cannot be identified without unreasonable effort, does not stand in the way of the general conclusion that the combination of a MAC address of a WiFi access point with its calculated location, should be treated as personal data. Under these circumstances and taking into account that it is unlikely that the data controller is able to distinguish between those cases where the owner of the WiFi access point is identifiable and those that he/she is not, the data controller should treat all data about WiFi routers as personal data*".[18] The Art. 29 WP explicitly acknowledges that more and more data might lead to identifiability as: *"people tend to disclose more and more personal location data on the Internet, for example by publishing the location of their house or work in combination with other identifying data. Such disclosure can also happen without their knowledge, when they are being geotagged by other people. This development makes it easier to link a location or behavioural pattern to a specific individual"*.[19]

2.5 Proposal Data Protection Regulation

In January 2012 the European Commission presented a Proposal for a General Data Protection Regulation.[20] As the proposed changes are stipulated as radical by many

[16] WP 136, WP 185, and Opinion 01/2012 on the data protection reform proposals, 23 Maart 2012, WP 191, p. 9. Website: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2012/wp191_en.pdf..

[17] Article 9 Directive 2002/58/EC.

[18] Article 29 Data Protection Working Party, 2011. Opinion 13/2011 on Geolocation services on smart mobile devices, adopted on 16 May 2011 (WP 185), p. 11. Website: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2011/wp185_en.pdf

[19] WP 185, p. 10.

[20] Proposal for a Regulation of the EU 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

privacy lawyers[21], it is interesting to see if any changes have been proposed in view of the concepts of personal and location data. Analyzing the new definitions of *personal data* and *data subject* reveals only some reshuffling of the texts of article 2 and recital 26 of Directive 95/46/EC, without any real changes. There is however an odd sentence in the proposed Recital 24: *"(…) It follows that identification numbers, location data, online identifiers or other specific factors as such need not necessarily be considered as personal data in all circumstances"*. This clearly is a deviation from the previous interpretations given by the Art. 29 WP. In a recent opinion - regarding the data protection reform proposals - the Art. 29 WP has already advised to change this too narrow interpretation of the concept of personal data.

Field Code Changed

2.7 Personal location data in real life

From this legal analysis of the concept of personal data it becomes clear that - especially when considering things like the Internet, social media and smart phones – we should be conscious about all the information we process in any kind of way. Because of the broad interpretation, chances are high we are dealing with personal - or worse location - data to which a whole array of legal rules apply. Before discussing whether this trend of expansion will actually strengthen privacy and data protection, the next section will illustrate from a technical perspective how quickly trivial data can become personal data.

3. **Identifiability of the "De-Identified" Data: A Technical Perspective**

3.1 From trivial to identifiable, some examples

Type and scale of data collected about people is ever increasing due to developments in technology and new applications such as social networking and real-time data sharing. Type and complexity of the data may vary, but the problem of identifiability of the "de-identified" data remains the same. In this section, we are going to give an overview of the research results showing the identifiability of various data types. Lets first consider simple tabular data, where rows correspond to individuals and columns correspond to attributes of these individuals. The sample tabular data provided in the figure below[22] contains health information together with some demographics of the patients.

(General Data Protection Regulation) Brussels, 25.1.2012 COM(2012) 11 final, 2012/0011 (COD). Website: http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf

[21] See for example the websites of http://www.osborneclarke.co.uk and http://www.allenovery.com/

[22] Pierangela Samarati. "Protecting Respondents' Identities in Microdata Release." IEEE Trans. Knowl. Data Eng. 13(6): 1010-1027 (2001).

Medical Data Released as Anonymous

SSN	Name	Race	DateOfBirth	Sex	ZIP	Marital Status	HealthProblem
		asian	09/27/64	female	94139	divorced	hypertension
		asian	09/30/64	female	94139	divorced	obesity
		asian	04/18/64	male	94139	married	chest pain
		asian	04/15/64	male	94139	married	obesity
		black	03/13/63	male	94138	married	hypertension
		black	03/18/63	male	94138	married	shortness of breath
		black	09/13/64	female	94141	married	shortness of breath
		black	09/07/64	female	94141	married	obesity
		white	05/14/61	male	94138	single	chest pain
		white	05/08/61	male	94138	single	obesity
		white	09/15/61	female	94142	widow	shortness of breath

Voter List

Name	Address	City	ZIP	DOB	Sex	Party	.....
.....	.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....	.....
Sue J. Carlson	900 Market St.	San Francisco	*94142*	*9/15/61*	*female*	democrat	.....
.....	.....	.....	.....	.....	.....	.....	.....

In the past, removing personal identifiers from data was considered enough for anonymization which was proven wrong by Samarati and Sweeney in 1998[23][24]. For example in the table above, the category *Health Problem* contains sensitive information, and should not be released with the personal identifier, therefore the *SSN* and *Name* are blinded. Only the *Race*, *DateOfBirth*, and *Sex*, together with the *ZIP* and *Marital Status* have been preserved because these attributes are useful for research purposes such as finding the correlation between health conditions and demographics or location. The attributes, *DateOfBirth*, *Sex*, and *ZIP* are not direct identifiers, but when they are used in combination, someone can link a public table with identifiers to a private table with sensitive information. For example in the table above, we see that the voters list does not contain private information, therefore releasing it should not be problematic, however, one can link the sensitive health information with the personal identifiers in the voters list through the birth date, zipcode, and sex which act like an identifier. In fact, Sweeney later on showed that through a very striking example of reaching the personal health records of the major of Massachusetts by linking his birth date, zipcode, and gender attributes in the supposedly anonymous health records with a public database.[25] This showed that attributes like zipcode, birthdate, gender are not identifiers but they could still be used to link to other identifier information stored in public databases, and therefore they need to be treated as quasi-identifiers.

[23] Pierangela Samarati, Latanya Sweeney: Generalizing Data to Provide Anonymity when Disclosing Information (Abstract). PODS 1998: 188

[24] L. Sweeney. k-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), 2002; 557-570.

[25] L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

The problem of re-identification of tabular data by linking with other data sources was demonstrated long time ago (and solutions were proposed which will be discussed in the next section), but similar re-identification leading to privacy leaks kept occurring for different data types. For example, in the AOL case[26], the data released was not tabular data, but search queries, and no further information was released except for the successive queries of people without any apparent identifier such as the IP addresses. The successive queries by the same user have been thought to be anonymous until some journalist was able to pinpoint an individual via her queries. This was possible since people search for their friends, things in their vicinity. Some people even search their names to look for things published about them on the internet or to see if they are visible on the internet. Such queries on the web reflect our age, sex, and location, and they act as quasi identifiers like it was shown tabular data.

In the case of social network applications, the problem is aggravated since there are all kinds of textual information about people plus their friendship information, and whatever their friends tell about themselves. For example, two MIT students (now graduates) Carter Jernigan and Behram Mistree analyzed the gender and sexuality of a person's friends to predict that person's sexual orientation, using a software program they developed[27]. It was not possible to estimate the accuracy of the program but through experimenting among their classmates, they found that the program accurately identified the sexual orientation of male users by analyzing the characteristics of their friends within their social network.

3.2 Location data

Location-based services have been in use for some time but with large companies such as Google promoting its location service Google Latitude, it has become a concern of privacy. Even though the law does not qualify location data as sensitive data, the nature of these data can be sensitive to a large extent., e.g. indicating presence in a hospital or a red light district. In addition it can be used to identify the person being in the hospital or the red light district. From this perspective the Art. 29 WP interpretation that location data are personal data is correct. Because of its sensitive nature, a case could even be made to qualify them as sensitive data in the sense of art. 8 of Directive 95/46/EC.

Even a simple Facebook status update to indicate the general location of the user, whether he/she is home or not, could be used by thieves which was the main idea of the sarcastic "pleaserobme.com" application which indicates the problems of revealing location data.

[26] See for a description of this case http://elliottback.com/wp/aol-gate-search-query-data-scandal/

[27] The New York Times. How Privacy Vanishes Online. By Steve Lohr Published: March 16, 2010.

It is not clear how much and how detailed location information is collected and stored by mobile service providers. For example a German Green party politician, Malte Spitz, discovered that we are being tracked voluntarily or non-voluntarily by cell-phone companies.[28] Mr. Spitz had to go to court to find out what his service provider, Deutsche Telekom, stored concerning his location. It turned out that in a six-month period — from Aug 31, 2009, to Feb. 28, 2010, Deutsche Telekom had recorded and saved his longitude and latitude coordinates more than 35,000 times. Mr Spitz wanted to show the privacy implications of this data and decided to release all the location information in a publicly accessible Google Document, and worked with Zeit Online, a sister publication of a German newspaper, Die Zeit, to map those coordinates over time. The visualization showed that Mr Spitz spent most of his time in his neighborhood and not much walking around. The data also showed that he flies sometimes when he could have preferred the more fuel-efficient train, an interesting detail for a Green Party member.[29]

With smart phones, the situation has become much worse in terms of what has been collected. A recent study has shown that most of the mobile apps are transferring location data together with identifiers without the user knowing it. For example according to a research conducted by WSJ, *"An examination of 101 popular smartphone "apps"—games and other software applications for iPhone and Android phones—showed that 56 transmitted the phone's unique device ID to other companies without users' awareness or consent. Forty-seven apps transmitted the phone's location in some way. Five sent age, gender and other personal details to outsiders."*[30]

Since location data can be very sensitive, we need to de-identify it before it is released. Initially one can argue that we can release location data after removing the directly identifying information such as the id, name etc. In fact, a single location without an identifier may not tell much. However, things may be different when geo-location data is collected together with a timestamp for a long period. Looking at the stops and moves of a person, one can identify the time spent in various places. For example, we can see that a person spends some time in a hospital rather than passing by, visits certain parts of the city, or goes to a mosque or church periodically.[31] Through some geo-visualization techniques and a detailed map of the environment, one can easily obtain a lot of sensitive information about the person whose trajectory is released. When there is a group of trajectories, one can also try to learn the relation-

[28] Noam Cohen, It's Tracking Your Every Move and You May Not Even Know, The New York Times, March 26, 2011.

[29] Idem.

[30] http://online.wsj.com/article/SB10001424052748704694004576020083703574602.html. Your Apps Are Watching You. By SCOTT THURM and YUKARI IWATANI KANE

[31] Mehmet Ercan Nergiz, Maurizio Atzori, Yücel Saygin, Baris Güç: Towards Trajectory Anonymization: a Generalization-Based Approach. Transactions on Data Privacy 2(1): 47-75 (2009)

ship of people from those trajectories by looking at the intersection points of the stops and moves.[32]

As we mentioned above, location data can be used to infer the identity of a person even without an explicit identifier attached to it. Again by looking at the stops and moves in the trajectory at certain time intervals, we can speculate that a person lives at a specific location if the person stops at that location at night most of the time, and we can also infer that a person works at a certain location or studies at a certain location if the person stops there and spends most of the day. We can do a simple address search to see who is living at that location or who is working at a specific location to link a "de-identified" trajectory to an individual.

In some data mining applications it is enough to release pair-wise distances among data objects. However, research results showed that with some background information, we can recover the exact values from the distances[33]. In order to prove our point, lets consider a very simple data set, provided in the table below taken from one of our previous research papers[34], where we just release the distances between the ages of people instead of the exact ages. We have 5 people, X1, through X5, and the distances among those people are just the difference of their ages, for example the distance between X2 and X3 is 91 through we do not know their ages. Now consider that an adversary knows the ages of two people among them, say X1 and X2, say 20, and 90. From that information, the adversary can discover all the rest of the ages, for example, knowing the age of X1 as 20, the age of X3 can be either 1 or 41 since its distance to 20 is 1. Knowing the age of X2, as 90, the age of X3 can be either 1 or 111. So the two pieces of evidence when combined, we can conclude that the age of X3 is 1. Even without knowing the ages of those two individuals (X1, and X2), we can still recover the ages of other people. For example age difference between X2 and X3 is 91, which is the maximum distance, meaning that these are the youngest and oldest people in the community. We can assign the minimum age 0, to X2 to start with, and then X3 will be 91. With that assumption, we can get an initial estimate of the ages of the rest of the population, and see if the estimate fits the known distribution of the ages, we can shift the estimate one by one until the distribution of the estimate matches the distribution of the ages in the society.[33][34] This simple method was shown to work on data objects with multiple dimensions as well such as trajectories, which are sequences of time-stamped locations.[35]

Formatted: Superscript

[32] MS Thesis. Ercument Cicek. Ensuring location diversity in privacy preserving spatio-temporal data mining (Sabanci University, 2009).

[33] E. Onur Turgay, Thomas Brochmann Pedersen, Yücel Saygin, Erkay Savas, Albert Levi: Disclosure Risks of Distance Preserving Data Transformations. SSDBM 2008: 79-94.

[34] Idem.

[35] Emre Kaplan, Thomas Brochmann Pedersen, Erkay Savas, Yücel Saygin: Privacy Risks in Trajectory Data Publishing: Reconstructing Private Trajectories from Continuous Properties. KES (2) 2008: 642-649.

	X_1	X_2	X_3	X_5	X_5
X_1	0	70	21	35	7
X_2		0	91	35	63
X_3			0	56	28
X_4				0	28
X_5					0

In the case of trajectory data, there is an unlimited amount of background knowledge that could be used for inferencing. For example in the popular iPhone App case, the nearest wi-fi locations were kept for a given user.[36] Using this information not only the wi-fi location could be found but also it could be used to pinpoint individuals. The same idea that is described above for discovering private information via distances can be used to find the location of people with a reasonable accuracy.[37]

3.3 Some Technical Solutions

Privacy in the context of location based services has been studied, and some solutions have been proposed to protect the privacy of individuals.[38] These solutions try to limit the accuracy of the location data that is sent to the service provider, and also try to break the link between the successive locations of the same individual, in a way limiting the ability of the service provider to reconstruct the trajectory of the individual. In case of static trajectory data release, anonymization techniques have been provided, where the main idea is to release generalized locations, as in the case of tabular data. This way we make sure that there are at least k people with the same trajectory, in a way hiding the people within crowds. However, this has its own limitations, because the sensitive locations visited are not considered in such anonymization techniques. For example, we may say that at least k people have stopped at a certain location and we can not distinguish them from each other, however if the stopped place is a sensitive location such as a hospital specialized in cancer treatment, then we know that all those people may have cancer.

Although these solutions have some limitations, they can still be enhanced and adopted for location data. In case of Location Based Services, existing solutions for enhanced privacy may be deployed by service providers. For example in order to respond to a service request, the exact location of the user may not be needed, and successive location information leading to the reconstruction of the trajectory of the user may not be necessary. However, the companies are reluctant to adopt these solutions and they prefer to rely on the consent of their customers to resolve privacy issues. There may be various reasons as to why privacy enhancing technologies for location data is not being widely used. One of these reasons could be that not all privacy en-

[36] http://technews.am/conversations/boy-genius-report/apple_sued_over_iphone_location_tracking_scandal

[37] Emre Kaplan, Thomas Brochmann Pedersen, Erkay Savas, Yücel Saygin: Discovering private trajectories using background information. Data Knowl. Eng. 69(7): 723-736 (2010).

[38] Chi-Yin Chow, Mohamed F. Mokbel: Trajectory privacy in location-based services and data publication. SIGKDD Explorations 13(1): 19-29 (2011).

hancing techniques are mature enough to be deployed[39]. But, the main reason is that, consent is an easy solution for the companies since they do not need to implement the privacy enhancing solutions which means extra cost for them. Users are also not given much choice but to accept the consent or not use the service.

4. **Conclusion**

On the basis of the above we can speak of an imbalance in the technological evolution regarding data processing. On the one hand, technologies enabling identification are flourishing. These technologies have created a situation in which enormous amounts of – at first glance trivial - data can be linked to a person, bringing these data within the scope of data protection regulation. This evolution is even magnified by the extensive interpretation given to the concept of personal data, including all location data. As demonstrated by the cases presented in section 4, sequences of locations belonging to an individual can easily provide evidence as to who that person is, where (s)he has been and with whom. On the other hand, technologies de-identifying data do not reach their aim in practice because of constantly improving linking and matching technologies and of the enormous amount of data sets (publicly) available.

It is interesting to link this technological conclusion to the goals of the EU data protection regime: *"the protection of individuals with regard to the processing of personal data"* and *"the free movement of such data"*[40]. Instead of contributing to these goals, the result of extensive interpretation of the concept of personal could have the opposite effect. Expanding the applicability of the EU data protection regime to daily processing activities and trivial data will decrease awareness for the need to comply with data protection regulations. This need is felt when data processing infringes upon private life, but with the extensive application of data protection legislation the link with privacy, the fundamental human right in which data protection finds its origin, seems completely lost. Moreover, in view of the second goal of Directive 95/46/EC, it seems that extensive application of the legal regime will rather hamper than improve the free movement of data.

[39] P. Ohm, Broken Promises Of Privacy: Responding to the Surprising Failure of Anonymization, UCLA Law Review nr.57, 2010, p.1701-1777

[40] These goals are described in the title of Directive 95/46/EC.