Deliverable D6.4:
LEGAL OPINIONS

Mobility Data Mining and Privacy
MODAP

Sabanci University
CNR - Area Della Ricerca di Pisa
University of Milan

Project reference: INFSO-ICT-245410

	LEGAL OPINIONS **Del. no: D6.4**	
INFSO-ICT-245410	**Version: 2.0**	**30/03/2013**

Deliverable type: R

Classification: **RE**

Work package and task: WP6, T6.1

Responsibility: **SU**

Executive summary:

This deliverable is the collection of opinion reports written by the members of the Privacy Observatory WG. Opinion reports are co-authored by experts in Law, however in some of the reports the reviewers will notice that the author is just indicated as "A Lawyer from the Privacy Observatory". This is due to the fact that the corresponding authors have a conflict of interest due to their clients. The author name is public to the Commission, but can not be disclosed in the published document. Three of the opinion reports were extended an published as papers. For these three opinion reports we included the versions as papers at the end of this document.

Contents

1. Introduction

In the context of MODAP, Privacy Observatory working group have been producing opinion reports on various privacy issues. Some of these reports are country specific and some are generic. We have chosen the Turkish privacy landscape since there are many ongoing e-government projects especially on health and social security, while the privacy laws are yet to be passed from the parliament. Therefore, there is a lot of potential for local impact in Turkey that MODAP tried to achieve through the experience of its members from European Union and other countries with established privacy laws. We have generated draft opinion reports provided in the following sections which are open for public discussion.

Each opinion report is summarized below with some information about its context and its context.

- The first opinion report with title "ROCKY LANDSCAPE OF THE TURKISH REGULATIONS ON PRIVACY OF PERSONAL DATA" is written by Erkay Savaş and Tuğrul Sevim. Erkay Savaş is a professor of Computer Science specialized on computer security and privacy, and Tuğrul Sevim is a lawyer specialized on data protection, and he is a lecturer at Bilgi University in Istanbul. This opinion report gives a (recent) historical background on the privacy laws in Turkey and summarizes the current state. It can be considered as an insider view of the current state in Turkey.

- The second opinion report with title "PRIVACY AND DATA PROTECTION IN TURKEY: (INCHING) TOWARDS A EUROPEAN FRAMEWORK" is written by Omer Tene and Yucel Saygın. Omer Tene is a professor of law from Israel, and Yucel Saygın is a professor of Computer Science. We are really thankful to Omer Tene for visiting Istanbul for working on this report during the height of the Crisis between Turkey and Israel. This report, like the first one, looks at the current state in Turkey in terms of data protection but including an outsider's view. We should note that Israel recently had negotiations with Europe for privacy regulations and therefore Israel and Turkey can share their experience in this path. This report was accepted for publication in the BNA's ***Privacy & Security Law Report*** which provides weekly coverage of the latest legal, regulatory, legislative, and judicial news in the privacy and security fields. The BNA report is attached at the end of this deliverable.

- The third opinion report with title "DRUG TRACKING SYSTEM AND PRIVACY ISSUES IN TURKEY: A CASE STUDY", was written by Roberto Lattanzi, Marilena Missorici, Dilek Tapucu and Burak Galip Aslan. Roberto Lattanzi and Marilena Missorici are from the Data Protection Authority of Italy. Roberto Lattanzi was seconded expert on Data Protection in the European Commission until he joined the Italian Data Protection Authority. Dilek Tapucu is a Computer Scientist and Burak Galip Aslan is also a Computer Scientist specialized on ethics. This report gives a closer look on the privacy problem of a

specific e-Health system in Turkey and provides some solutions in European dimension.

- The fourth opinion report with title "PIVACY BY DESIGN IMPLEMENTATION AND GUIDELINES FOR THE E-GOVERNMENT PROJECTS IN TURKEY" was written by Prodromos Tsiavos who is from University College London, School of Law, and also advisor to the Greek authorities on e-Government projects. The author provides guidelines for the e-government projects in Turkey in terms of privacy by design.

- The fifth opinion report with title "PRIVATE PLACES: PROTECTING PRIVACY AGAINST LOCATION PROVIDERS" is by Omer Tene and Maria Luisa Damiani. Maria Luisa Damiani is a professor of Computer Science and in this report the authors identify a serious privacy problem in terms of possible pervasive location tracking by service providers. Omer Tene would like to stay as an anonymous author due to conflict of interest with his clients. This opinion report has been revised, extended, and published in a conference. We include the extended version of this opinion report at the end of this deliverable. The title of the extended/revised version is "Location privacy challenges in geolocation standards: an interdisciplinary study"

- The sixth opinion report with title "FairKDD" was written by a group of experts in data mining, privacy preserving data mining, and law. This report looks at a broader problem of data mining, and proposes a roadmap for the fair practices for knowledge discovery and data mining. It is significant number of experts from different parts of the world indicated their support by signing this document.

- The seventh opinion report with title "THE RISE OF IDENTIFIABILITY: THE DOWNFALL OF PERSONAL DATA?" is written by Colette Kuijpers and Yucel Saygın. Colette Kuijpers is a professor of Law from TILT, Netherlands. This report can be considered as work in progress, and it was accepted for presentation at CPDP Conference in Brussels in January 2012. The extended version of this work is attached at the end of this report with the same title.

- The eighth opinion report is about an ongoing project of Turkish Telecom which is about building a recommendation system about users taking into account their past behavior. Since this project is about sensitive data, Erkay Savaş looked at the system and the possible privacy problems.

- The ninth opinion report is a about a serious conflict with the privacy settings of Facebook and the querying facilities open to public. This was written by Andrea Zanda (A PhD Student from UPM, Madrid, who did the technical study), Omer Tene and Yucel Saygın. Omer Tene would like to stay anonymous in this report due to conflict of interest with the clients. We have reported this to Facebook discussion board, and now Facebook has changed its system to avoid this privacy problem.

- Finally the the tenth opinion report is on the open data and privacy dilemma which describes the challenges of open data from a privacy perspective and provides some hints as to how they could survive together.

2. Rocky landscape of the Turkish regulations on privacy of personal data

Authors: Erkay Savaş and Tuğrul Sevim

Privacy as a Constitutional Right

Article 20 of the Turkish Constitution identifies the individual privacy as a constitutional right by stating, "Everyone has the right to demand respect for his private and family life. Privacy of individual and family life cannot be violated." In addition, the third ammendment to the Turkish consititution dated 2010, which recognizes the protection of personal data as constitutional right, necessitates a law to define the principles and procedures. This ammendment under the gives the individuals the right to demand the protection of their personal data. More specifically with this right, the individuals are entitled to demand to be informed about the data collected about them, to check if these collected data are being used for the specified purpose, to access to these data and demand that they are modified or deleted. Also by the ammendment, personal data will only be processed with the consent of the data owner, or in cases provisioned by the law.

Current Status of the Law for Privacy of Personal Data

Turkey was a signatory of the "Convention 108" for the Protection of Individuals with regard to Automatic Processing of Personal Data, which was opened for signature by the Europeran Commission on 28 January 1981. The Convention was effective as of 1985 and ammended in 1999. Although a signatory of the Convention, Turkey didn't ratify it to transpose it to its internal judicial system. Turkey is among the three countries that did not ratify it (the other two countries are Russia and Armenia). The convention stipulates that a signatory nation pass a law to regulate the protection of personal data. However, currently, there is no such law in Turkey.

As being a candidate country, the fact that Turkey doesn't have a compliant law was also criticized in progress reports on Turkey's accession to European Union. The latest Accession Partnership document dated 2008, where the specific actions are planned to be taken by the relevant national institutions to adres the issues in the progress report, mentions the priorities regarding the protection of personal data. Two priorities that need to be addressed in one or two years are:

- Bringing the Turkish legislation in comply with "the Community acqui" of the European Union for the protection of personal data

- Increasing the efforts to establish data protection agency which are similar to European counterparts

These two priorities neccesitate legal regulations in comply with "the Community acqui" of the European Union.

Note also that the ammendment to Article 20 regarding the protection of personal data also necessitates a specific law to regulate the privacy of personal data. But, currently there is no current law to define the principles and procedures needed for the protection of personal data. On the otherhand, there has been a draft law for the "Protection of Personal Data" (Draft Data Protection Act (DDPA)) since 2003. It has actually been sent to the relevant legislative commission at the Parliament , but not yet been adopted. The DDPA is publicly accessible. The draft is not in its final shape and efforts to give the final form of the draft are curretnly ongoing. Since the subject matter of the draft law is of interest to many public institutions, there is an incentive to accomodate their concerns in the law, which may be one of the resons of its delay. Although there are rumours that the draft will become a law either in late 2011 or 2012, it seems to be impossible to give an exact time schedule as to when this law has been passed in the parliament.

When one inspects the draft law that is now with the Justice sub-committee of the Turkish parliament, one can clearly see that it basically provides a general framework for the protection of personal data. The draft law apparently is based on the European Union Directive with the code number 95/46/EC on the protection and regulation of personal data. In this draft that the definitions, the fundamental principles on the processing of personal data, the rule that necessitates the consent of data owner to process the personal data and the exceptions to this rule, rights of the data owner, the issue of forming the Personal Data Protection Agency or Body clearly indicates that the general framework that the draft law is providing is very similar to the framework of the European Directiveof 95/46/EC. The draft law being in comply with the European Directive is important per se. On the other hand, the European directive itself was prepared 20 years ago. Technological progress that increases the means and volumes of collecting personal data brings about new challenges. There is a common understanding that the framework in the original directive is not sufficient to address the new challenges. The European Union has acknowledged this situtation and decided to renew the existing framework for the data protection in the context of European Digital Agenda. The efforts on this area is stil continuing. Turkey needs to closely follow these developments in order to both have a law that is in comply with European directives and enable to meet the challenges posed by new application areas and new technological developments.

There is currently a working group organazided by the Ministry of Justice to bring the DDPA up to date with the current developments and the needs of the technology. The next workshop will be held on October 7,2011, an our group will participate in the workshop.

Some Expressed Concerns About Draft Data Protection Act

The DDPA prohibits the processing of the personal data (e.g. race, political views, religion or affiliation of other faiths, membership to organizations, medikal data and

convictions). However, it also states the exceptions when these personal data can be processed provided that sufficient countermeasures for the protection of personal and family life are guaranteed. One particular example is that medical data can be processed to provide better e-health services by many organizations (e.g. medical service providers, insurance companies, social security,medicak schools and universities) under the supervision of professionals who are under oath to protect the patients privacy. By this particular statement, some experts claim that the law basically leave the regulations on medical data to yet another law since these statements are not sufficently clear to provide implementable regulations.

Turkish Ciriminal Code

The 2005 Criminal Code regulates crimes regarding the private life. Storage, illegal transfer or retention of data are considered as offences.

Sectoral Regulations

In some sectors in Turkey, the regulations are considered to be satisfactory. One such sector is telecommunication sector. The Telecommunications Council (BTK in Turkish) is the institution responsible for data protection and privacy issues in the telecomunications sector. A regulation was enacted in 2004 by the Parliament for data protection in the telecommunication sector. The Regulation is in principle a summary of the European Union's Directive 2002/58/EC on data protection in electronic communications. The Act prohibits collection of personal data without the data subject's express consent and processing of the personal data by means such as telephone, fax, mobile phone, and e-mail.

Conclusion

The fact that Turkey does not have yet passed a law for the protection of personal data is not acceptable as the need and problems are intensifying. However, there is a considerable effort to complete this procedure to provide an acceptable level of regulation. However, even if the law was put into the effect, there will be other concerns. One such concern is that the law does not sufficiently regulate some sectors such as medical data and necessitates another law for this purpose. Also, this is an indication that the regulations particular to sectors will likely to play an important role. Another concern is that the definition of personal data is not provided except for the regulations in telecommunications sector where certain types of data are considered to be private and cannot be disclosed. One other issue that is voiced by some experts is that the Data Protection agency, when it is established, will not be totally independent but rather act under the auspices of the government. Practice will show whether this is indeed a negative aspect since some countries have similar situation where the Data Protection agency is not independent. Lastly, the working group for updating DDPA does not contain technical experts who can inform about the techniques to maximize the utilization of aggregated personal data without compromising privacy of personal data. Without technical expertise, the regulations can be too restricted to a degree that data utilization is minimized. Our participation in the working group aims to bring this expertise to the said efforts.

3. Privacy and data protection in Turkey: (Inching) Towards a European framework

Authors: Omer Tene[1] *and Yucel Saygin*[2]

Introduction

With a population of 73 million, second only to Germany among European Union Member States, and a GDP growth rate of 8%, Turkey is a force to be reckoned with on the outskirts of Europe. Turkey harbors the largest city in Europe (Istanbul, with an estimated population of 15 million); has 63 million mobile telephone subscribers; and is the fifth largest country in the world in terms of the number of Facebook users (after the United States, Indonesia, India and the United Kingdom) with 30 million.

Yet despite many years of political and public debate, Turkey remains without a comprehensive legal framework for privacy and data protection. Besides constituting an obstacle in Turkey's quest to become a member of the European Union, the lack of a modern data protection law limits business opportunities and potential collaboration between Turkish and foreign businesses and law enforcement entities and compromises the fundamental rights of Turkish individuals.

Constitutional

In March 2010, the Turkish government submitted to parliament a package of constitutional amendments, including an amendment to the privacy protection provision, Article 20, adding a section on data protection.[3] The constitutional amendments were authorized by the Turkish Grand National Assembly in May 2010, yet failed to garner the support of a two-thirds majority required to pass directly; gaining 336 of the 500 members of parliament, short of the 367 threshold. They were therefore submitted to a nationwide referendum in September 2010, after withstanding constitutional challenge in the Turkish Constitutional Court, where they passed with sweeping support of 58% of the vote.

Pursuant to the constitutional amendment, Article 20 now states that "everyone shall have the right to the protection of their personal data".[4] It provides that personal data may only be collected, stored, and used pursuant to a legal obligation or with the data subject's

[1] Managing Director, Tene & Associates; Associate Professor, Israeli College of Management School of Law.

[2] Professor, Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul Turkey. Both authors are participants in MODAP (Mobility, Data Mining, and Privacy), a Coordination Action project funded by the EU FET OPEN.

[3] The text added to Article 20 of the Turkish Constitution provides: "Everyone has the right to demand the protection of his personal data. This right comprises the right to be informed about the personal data concerning himself, access to such data, right to request correction or deletion of them as well as the right to be informed if such data is used in accordance with the purposes for which it was collected. Personal data can be processes only in cases regulated in a law and upon express consent of the subject individual. Principles and procedures regarding the protection of personal data shall be regulated by a law."

[4] Communications secrecy continues to be protected by Article 22 of the Constitution.

consent. It grants individuals the right to be informed about any processing of their personal data; as well as the right to access, rectify or delete their personal data. It calls for legislation setting forth the principles and procedures to implement the constitutional mandate. Yet even before such legislation is put in place, Turkish individuals benefit from constitutional protection *vis-à-vis* action by public sector bodies. Indeed, in its latest progress report, the European Commission states that "with regard to fundamental rights, progress has been made. Constitutional amendments bring important changes in the area of data protection (…)"[5]

Finally, Turkey is currently in the process of introducing a new constitution to replace the existing document, which dates from 1982, a relic of the 1980 *coup d'état*.[6] Given that privacy and data protection are not politically controversial, we hope that the new constitution will echo the provisions of the recently amended Article 20.

Omnibus legislation

Omnibus data protection legislation has been contemplated in Turkey for almost a decade, yet has been slow in coming. Turkey is a signatory to the Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data ("Convention 108"); yet has not ratified the Convention. It is has also been a member of the Organization for Economic Co-operation and Development (OECD) since 1961. As part of Turkey's 2005 Accession Partnership with the European Union, Turkey was required to "adopt a law on protection of personal data" and "establish an independent supervisory authority."[7]

According to the European Commission's 2009 progress report, Turkey is yet to take the required measures to implement this data protection agenda. The Commission states: "With regard to respect for private and family life and, in particular, the right to protection of personal data, Turkey needs to align its legislation with the data protection *acquis*, in particular Directive 95/46/EC, and, in that context, to set up a fully independent data protection supervisory authority. Turkey also needs to ratify both [Convention 108] and the additional protocol to it on supervisory authorities and trans-border data flow (CETS No 181). Wiretapping has reportedly been used extensively and records of it published in the press."[8] In its 2010 progress report, the Commission emphasizes that an "effective personal data protection regime is crucial for efficient international judicial cooperation."

---

[5] Commission Staff Working Document, Turkey 2010 Progress Report, accompanying the Communication from the Commission to the European Parliament and the Council, Enlargement Strategy and Main Challenges 2010-2011, 9 November 2010, http://ec.europa.eu/enlargement/pdf/key_documents/2010/package/tr_rapport_2010_en.pdf.

[6] Ece Toksabay, Turkish prime minister wants new constitution by first half of 2012, Al Arabiya, 29 September 2011, http://www.alarabiya.net/articles/2011/09/29/169270.html.

[7] Council Decision 2008/157/EC of 18 February 2008 on the principles, priorities and conditions contained in the Accession Partnership with the Republic of Turkey and repealing Decision 2006/35/EC.

[8] Commission Staff Working Document, Turkey 2010 Progress Report, accompanying the Communication from the Commission to the European Parliament and the Council, Enlargement Strategy and Main Challenges 2009-2010, 14 October 2009, http://ec.europa.eu/enlargement/pdf/key_documents/2009/tr_rapport_2009_en.pdf.

In 2003, the Turkish Ministry of Justice tabled a Draft Law on the Protection of Personal Data (the "Draft Law"); yet the Draft Law has since been lost in parliamentary committee hearings, while Turkey remains without proper legislation. The Draft Law is predicated on the European Data Protection Directive (95/46/EC). It applies "to natural and legal persons whose personal data are processed" and defines personal data as "any information relating to an identified or identifiable natural or legal person." Hence, unlike most European data protection laws, the Draft Law would protect both individuals and legal entities. Similar to European data protection legislation, it applies to both automated and manual data processing. The Draft Law sets forth principles for processing of personal data, including purpose specification, proportionality, accuracy, and retention limitation. It lists legitimate bases for processing personal data, including consent and compliance with a legal obligation.

Article 7 of the Draft Law, which applies to sensitive data, permits processing "[i]f the personal data are required to be processed for purposes of preventive medicine, medical diagnosis, medical treatment, health care or execution of health services by; (1) Health facilities, (2) Insurance companies, (3) Social security institutions, (4) Employers with the liability of establishing health unit at the work place, (5) Schools and universities; in accordance with the relevant laws, under the supervision of health care personnel (…)". We believe that this provision, which is based on Article 8(3) of the European Data Protection Directive, is overly broad. It exposes highly sensitive data not only to health care providers but also to insurance companies, social security institutions, employers and schools; indeed, it is difficult to foresee *any* use of medical data that is not authorized by Article 7. As such, it lacks adequate safeguards to protect the privacy of patients and their families. Hence, even if it is pushed through the legislative process, the Draft Law remains deficient.

Other provisions in the Draft Law appear internally inconsistent. For example, under Article 5(2) of the Draft Law, "[p]ersonal data can be re-processed for scientific or statistical purposes or for a longer duration than stated (…), provided that there is adequate safeguards in the relevant legislation regarding the purpose of re-processing (…)". Thus, this provision allows processing of data for secondary purposes, including scientific or statistical purposes, without expressly requiring that such data be anonymized. At the same time, Article 10 of the Draft Law provides that "[f]or research, planning and statistical reasons which aim at public benefit, the personal data can be processed provided that they are made anonymous".
Additional provisions of the Draft Law provide individuals with rights of transparency, access and rectification (Articles 11-13); and impose data security obligations (Article 15); as well as restrictions on crossborder data transfers based on the principle of adequacy (Article 14). Certain provisions of the Draft Law already appear outdated, such as those setting up a registry of databases (Articles 16-19). Even in Europe, the bedrock of the database registration (notification) regime, regulators have come to accept that in a globalized economy with ubiquitous data flows, such bureaucratic exercises yield few tangible benefits. The Draft Law further sets up a data protection authority as well as an independent data protection board (Articles 26-33); although the relation between the two

is not entirely clarified by the legislative text. Finally, the Draft Law sets forth penal, administrative and civil causes of action for willful or negligent unlawful processing of data.

Almost a decade has passed since the introduction of the Draft Law – a very long time in terms of technological progress and business development. Innovations and breakthroughs, particularly in information and communications technologies, have created new business models and tools affecting individuals' lives and impacting the functioning of virtually every businesses process and government activity. This means that the Draft Law is in danger of becoming obsolete even before its enactment. Indeed, the European Data Protection Directive, which serves as its main model, is subject to comprehensive review by the European Commission this year. In devising its data protection legislative strategy, Turkey should take account of these transformations to avoid adopting an outdated scheme.

Sector-specific regulation

With omnibus data protection legislation stalled and the establishment of a data protection authority delayed, Turkey has begun developing sector-specific regulatory schemes. First and foremost is the framework for protection of privacy and personal data in electronic communications. In 2000, Turkey established the Telecommunications Council as the institution responsible for data protection in the telecoms sector. The Council was responsible for the enactment in 2004 of the By-Law on Personal Data Processing and Protection of Privacy in the Telecommunications Sector (the "Regulation"); as well as the 2008 By-Law on Security of Electronic Communications. The Regulation is based on the principles of the European e-Privacy Directive (2002/58/EC), such as communication confidentiality and security; the handling of traffic and location data; call number identification; public subscriber directories; and prevention of spam.

In the past few years, the Council now named the Information and Communications Technology Authority ("ICTA") has enforced the Regulation forcefully, imposing steep penalties and making its presence felt in the telecom sector. In one case, *Taraf*, a Turkish newspaper, disclosed that personal data of subscribers of a telecom operator could be accessed by third parties. After a thorough investigation, ICTA issued the operator a fine in an amount of 1.250.000 TL ($800.000). Another case involved illicit access by a former Turkish soccer star, Ridvan Dilmen, to call traffic records of his ex-girlfriend. An investigation launched by the ICTA and carried out by the police ultimately led to the imposition of a fine in an amount of 13 million TL ($9 million).

Additional sector specific legislation applies in the financial sector, under the Banking Law of 2005, which imposes a strict duty of confidentiality subject to criminal and civil sanctions; and the Bank Cards and Credit Cards Law of 2006, which restricts retention or re-use by retailers of customer information as a result of credit transactions.

Alas, at the same time, many sectors of the Turkish economy remain largely unregulated, including important government projects such as electronic health records, which contain

highly sensitive information about patients' prescription histories and lack adequate measures of privacy and data protection. For example, the Turkish Ministry of Health established a comprehensive Medication Tracking System ("MTS") for purposes such as ensuring drug safety, accounting for use of medications and engaging in pharmaco vigilance. The MTS is based on a Ministry of Health regulation titled "Medicine Tracking and Evaluation". Under this scheme, massive amounts of personal data, including name, age, diagnoses, hospital visited, medicines purchased, and adverse effects, are collected and stored in a centralized database. Not only health care practitioners but also pharmacists are able to query the database about a patient's prescription history, as well as additional information such as address and social security payments. Even more sensitive health data are collected and stored by Turkey's Social Security Administration.

The Medicine Tracking and Evaluation regulation fails to clearly articulate the purposes for collection of patients' data; roles and responsibilities of various actors involved; procedures for reporting adverse medicine effects to third parties; and scope of access authorizations. Moreover, the MTS appears to violate data protection principles such as proportionality and data minimization, collecting large amount of personal data and making them available to numerous parties without apparent necessity. It is evident that only comprehensive data protection legislation can quench the thirst of public sector entities for additional data, and subject such entities to requirements of data minimization, security, and retention limitation.

Moving ahead

In order to give substance to the latest constitutional amendments and progress toward harmonization and partnership with the European Union, Turkey should adopt data protection legislation based on the fair information principles ("FIPs") set forth in the 1980 OECD Privacy Guidelines. In addition, Turkey should establish a privacy and data protection authority charged with enforcing data protection legislation against both the private and public sector, in a manner commensurate with the structure of Turkey's constitutional and administrative law.

While procedures for implementation and enforcement diverge, there is broad global consensus concerning the substantive data protection principles, such as purpose limitation, proportionality, transparency, and security. Indeed, even the United States, which does not yet have cross the board data protection legislation, has recently endorsed regulation based on the FIPs. In its Report on "Commercial Data Privacy and Innovation in the Internet Economy: A Dynamic Policy Framework", the United States Department of Commerce states: "To provide consistent, comprehensible data privacy protection in new and established commercial contexts, we recommend that the United States Government recognize a full set of Fair Information Practice Principles as a foundation for commercial data privacy." Turkey, which has already recognized the importance of protecting the right to privacy and individuals' data in its constitutional amendment and acceded to Convention 108, should likewise adopt legislation setting forth the FIPs as statutory requirements. The FIPs should apply to both private and public sector entities and will be enforced by a dedicated data protection authority.

Turkey's data protection authority should be established in a manner commensurate with its constitutional and administrative law. Under a recent decree, Turkey subjected all independent regulatory bodies to the supervision of government ministries.[9] While the European Data Protection Directive mandates a supervisory authority acting "with complete independence", independence should be balanced against enforcement powers and accountability. Turkey's legal system favors strong enforcement agencies exercising broad powers and accountable to the judiciary branch over weak independent authorities effectively acting as ombudsmen. The European Commission approved the "adequacy" of data protection frameworks in Argentina and Israel, despite these countries' data protection authorities being integrated into the executive branch.[10] A similar structure, where the data protection authority is subject to strictly financial and administrative oversight of the Ministry of Justice, should be considered in Turkey.

Conclusion

Despite constitutional amendments and the establishment of a strong regulator in the field of telecom privacy, Turkey continues to lack a comprehensive data protection framework in the mold of European Union Member States. This imbalance, subjecting a specific sector of the economy to what has sometimes been considered heavy-handed regulation, while leaving other sectors without regulatory guidance or oversight, should be corrected. A data protection authority, established in the spirit of European regulatory agencies and empowered to enforce the law in all sectors of the economy as well as against the state, would correct this imbalance and propel the Turkish economy towards European standards.

[9] Independent bodies in Turkey attached to ministries after new decree law, Hürriyet Daily News, 22 August 2011, http://www.hurriyetdailynews.com/n.php?n=independent-bodies-in-turkey-attached-to-ministries-after-new-decree-law-2011-08-22.

[10] See, e.g., Commission Decision of 31 January 2011 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequate protection of personal data by the State of Israel with regard to automated processing of personal data, OJ L 27/39, 1 February 2011, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:027:0039:0042:EN:PDF.

4. Drug tracking system and privacy issues in turkey: a case study

Authors: Roberto Lattanzi, Marilena Missorici, Dilek Tapucu and Burak Galip Aslan

A (tentative) description of the DTS

The Ministy of Health has been empowered of "*ensuring drug safety*" by the by-law No. 181 about "The Duties and Establishment of Ministry of Health" (see clause 2.d.11.c which states that "*drug-tracking must be made possible*").

Accordingly, there had been a revision on "*Drug Packing and Coding Directive*" of Ministry of Health in 02.02.2008 which has led to the announcement No. 2009/84 (dated 31.12.2009) by Ministry of Health informing hospitals and medical centers to ensure all incoming drugs to be square-coded.

Following a directive from Ministry of Health about "*Drug Tracking and Evaluation*", an information system has been developed to track *all* the drugs purchased, identifying in the meantime the patient/purchaser. Pharmacovigilance is one of the various and not clearly defined purposes of this very vague directive; other mentioned purposes of the DTS system are fraud detection as well as two undefined purposes: namely the "*collection of data for the purpose of creating a policy*" and "*other benefits*" (see http://iegm.gov.tr/Default.aspx?sayfa=its_bilgi&lang=tr-TR).

Based on this directive, a large amount of personal data concerning (potentially) all Turkish citizens (name, surname, age, drugs purchased), are collected in a centralized database. In the next years, its content will be, day by day, enriched.

According to the directive "*the identity and address of health personnel and patient is kept hidden by Ministry of Health*". Without the data subjects' consent, the recorded information should never be disclosed to third parties except TUFAM personnel. Certified authorities, health institutions and health personnel are bound by the same privacy principles (see privacy clause 11).

Considering the "*ITS Usage Guideline*" published by the "*Drug and Pharmacy Department of Ministry of Health*", data security standard for user authentications are identified (see clause 12 addresses TSE ISO/EN 27001) and the authorized personnel can access database. Within the authorized personnel are mentioned ITS programmers, ITS helpesk personnel, "*data submitters*" and "*data queries*" (see clause 13).

But, as a matter of fact, the DTS as it is now designed:

a. makes it possible to the "health professionals" (doctors, pharmacists, dentists and nurses) to send information related to any purchase of drugs to the database (and not only in the cases to adverse effects);
b. allows pharmacists to query about the "drug history" of the patient (and other personal information: e.g. their address and social security debt status etc.). For instance, the pharmacists are empowered to see the drugs purchased by a given patient (independently from the appearance of any adverse effect) together with

the diagnosis and the hospital(s) visited. An authorized user can therefore abuse the system getting sensitive information and use it for other (not legitimate) purposes. For instance, insurance companies, employers or data subjects' relatives could be, in some cases, interested in the (drug usage and diagnosis) information contained in the database.

The DTS: An example of e-government application in the health sector. Some critical considerations (from a constitutional and international law perspective)

The DTS, as one of the multifarious e-government informational tools, merits to be carefully considered, as it is one of the first e-health applications introduced in the country and, in the meantime, it deals, at a very large scale, with very sensitive personal information; for sure, other ICT applications will follow in the next future.

Moreover, the DTS represents an excellent case study (other areas could be as well be taken into consideration) to investigate, more generally, the state of the art regarding the "privacy" topic in our country.

Unfortunately, as a starting point, it should be pinpointed that the DTS doesn't seem to take adequately into account the traditional duty of medical secrecy, as well as the modern privacy and data protection issues.

It is well known that worldwide – as well as in the Turkish legal system – health data are traditionally strongly protected via the doctor's confidentiality duties. The reason for this protection is, on the one hand, to guarantee the individual privacy and dignity, and, on the other hand, to protect (indirectly) the public health avoiding the spread of diseases.

Moreover, our Constitution – following the modern trends in western countries and the recent adoption of the European Charter of Fundamental Rights – recognizes now the new right to "data protection". According to the Art. 20.3 of our fundamental law *"Everyone has the right to demand the protection of his personal data. This right comprises the right to be informed about the personal data concerning himself, access to such data, right to request correction or deletion of them as well as the right to be informed if such data is used in accordance with the purposes for which it was collected. Personal data can be processes only in cases regulated in a law and upon express consent of the subject individual. Principles and procedures regarding the protection of personal data shall be regulated by a law."*

This Constitutional provision represents without doubt a first important step "in the right direction", but is only the first one. In fact, a general law on data protection – in conformity with the Council of Europe Convention n. 108/1981, signed (and not yet ratified) by the Turkey Republic (together, only, with Armenia and Russia) – is not yet in force.

Notwithstanding, even if an *omnibus law* – comparable with those existing in Europe, Australia and New Zealand, Canada and in some North-African and South-American

countries – has not yet been adopted, the clear choice of values and basic principles now enshrined in our Constitution cannot be neglected. These new principles should therefore be respected also in the case of the DTS – not yet fully in line with the new legal constitutional framework – and, in more general terms, they shall apply to any other e-government project.

Considering now the DTS (in the light of the directive of the Ministry of Health about "*Drug Tracking and Evaluation*") in a more detailed way, it seems to suffer from a number of weaknesses, and in particular a *lack of clarity* regards:

a. the purpose(s) of the collection,
b. the roles and responsibilities of the various actors involved,
c. the procedures for adverse drug reaction reporting: the kind of information on adverse effects which should be collected should be precisely indicated, as well as how it should be stored or how (and to whom) it should/could be communicated;
d. the subjects who can have access to the database (and the conditions under what the access can occur).

Clearly, all these aspects need to be fixed in the DTS case.

Moreover, personal data need to be used only if (and in the measure) they are necessary: identifiable health data shall only be processed when strictly necessary and parties involved should assess this necessity at every single stage of the pharmacovigilance process; on the contrary, in the DTS case, the collection of a large amount of personal data also in cases where apparently no adverse reaction takes place it is clearly disproportionate.

As said, the beginning of the "digitalizing era" of our health sector and the introduction of complex informational systems like the DTS – that in the future, as it is happening in other western countries, would benefit of the introduction of telemedicine and other tools to increase the health condition of the population (like the electronic health record system) – cannot avoid to be based on the existing constitutional framework (coherent with the international standards and with the ethical principles and values which are behind it), in order to respect the dignity, privacy and the data protection of the data subjects.

It is therefore desirable and advisable that the DTS system could be improved as soon as possible, in order to achieve legitimate purposes in the interest of public health while respecting the privacy of the individuals.

Any collection of personal data, especially of sensitive data such as health data, should be put in place with sound guarantees: privacy preserving techniques (such as anonymization or usage of pseudonyms) should be preferred; only data strictly necessary to the purpose established by a law (or in conformity with a law) can be collected; the data should be relevant to the purpose, correct and updated; the data subject (that is the person to which data belong) should be informed of the collection of his personal data, of the purpose of this collection, of the name of the data controller (that is the subject that are processing the data for the established purpose), of the other subjects to whom the data can be communicated. Data subjects should have the right to access to their personal

data, to correct them and to ask for their deletion if they are processed unlawfully (for example, if data collected are not relevant for the purpose for which they have been collected).

This communication of personal data from the database of the Ministry to pharmacists is not provided by the law and it is not necessary for the purpose of pharmacovigilance; clearly a free access to a wide amount of sensitive personal data can lead to strong abuses and violations of individual rights.

Law, institutional body & technology: An indispensable partnership to protect the person in the information (control) society

Data protection principles shall not be (or remain) written only in the "black letters" of law. Even if our legal system will be completed with the adoption of a data protection law – as we hope and warmly recommend –, in order to be fully respected and implemented, any legislative provision in the information society shall always be transposed in the (techno)logical architecture of the information system (i.e. in the hardware and software), possibly since the design phase of the informational infrastructure. In the north-american and European experience this need is now clearly expressed with the slogan "*privacy by design*".

Moreover, in order to correctly "design" any informational infrastructure, its possible impact on privacy rights need to be previously mapped. As a consequence, especially in large scale systems (as DTS), a "*privacy impact assessment*" should be carried out.

Auditing procedures, carefully targeted, should also be put in place to check, *ex post*, the lawfulness of the processing and the respect of security measures, in order to detect unlawful access to the database.

Fundamental rights – and among them, privacy and data protection – need to be effective. Besides individual judicial remedies, the enforcement of the data protection right should be granted by the establishment of an *independent administrative authority*. This body should – as some best practice in other countries demonstrate (see, e.g., the Italian example with the legislative decree n. 196/2003) – not only have sound enforcement powers, but also be in charge of giving opinions to the Parliament and the Government (if data protection issues are considered), raise public awareness among the citizens, evaluate if code of conduct adopted by the relevant stakeholders respect the data protection law.

How to speed up the process in Turkey?

In our view, the Academy has an important role in speed up the process of adoption of a legislative and institutional framework in line with the constitutional provision contained in article 20.

In this sense Sabanci University is approaching the data protection topic, especially (being a technological faculty) when ICTs are involved. Our participation in international projects (as MODAP: www.modap.org) and the organization of workshops, conferences

and etc) clearly confirm our commitment. We call for a wider cooperation in our country, especially with law faculties.

This is not enough. A sound institutional commitment is also needed.

International cooperation, especially with the European Union and the European member states – traditionally our major commercial partners –, is for sure a useful way to speed up the process. International cooperation should not remain simply an (abused) slogan.

Immediate steps could be taken, for instance, launching a "*twinning project*" in the data protection area, in order to facilitate the introduction of a national legislation that guarantees an adequate level of protection of personal data in Turkey – so that, with clear advantages for our industries, trans-border data flow with the European countries can take place without any obstacles – and the institution of a Turkish data protection authority.

This is our view. We will keep on working actively in this direction.

5. Privacy by design implementation and guidelines for the e-government projects in turkey: Cultivating regulation for supporting privacy by design

Author: Prodromos Tsiavos

Avgerinos Law Firm, Greece

Introduction

In this report we suggest two strategies in order to implement a Privacy by Design (PbD) approach in the protection of personal data. While PbD has been an initiative applauded by most Data Protection Authorities (DPAs), practitioners, academics and civil society active in the Data Protection area, it still remains a major problem how it is to be legally and administratively implemented. To appreciate how this is to be done, we may employ the Turkish drug monitoring and vigilance system as an example.

The problem

In order to be able to monitor the drugs quality, Turkish pharmacists have access to the whole of the medical history of the patient. There is no encryption or authentication system ensuring that the pharmacist or her assistant will not see or export personal patient information. While the ministerial decree specifying the rules under which the disclosure of such data should take place stipulates that this has to happen in a way that would respect the individual's personal data, there is no actual measures taken for implementing the relevant provision. The situation is further aggravated by the fact that while there is a relevant constitutional provision, there is no actual legislation addressing this issue.

Possible ways to tackle the issue

In order to have this problem resolved there are two different options:

- full legislative intervention
- intervention at the level of a technical standard

Each one of them requires special consideration. However, in both case we need to understand the structure of the regulatory environment to appreciate the merits of each type of intervention.

Levels of Regulation

In each legal system we may see a number of different regulatory instruments that in most cases have a hierarchical relationship with each other. These instruments are primarily legal but they are also administrative and technical. They form a pyramid of regulatory instruments which needs to be consistent with the basic data protection principles. Such pyramid has as a distinct feature, that while the higher forms of regulation are the most binding ones, the lower parts of the pyramid are the ones most frequently observed by the actors participating in the transaction.

At the top of the pyramid we have the top level international instruments such as the Convention for the Protection of Individuals with regards to Automatic Processing of

Personal Data, which sets a number of data protection principles. The second level is to see International regulations such as the EU Data Protection Directive (95/46/EC) that is then implemented in a further level that of national legislation. National constitutional rules are higher in the hierarchy compared to national laws and they are not necessarily in a hierarchical relationship with international treaties or regional regulatory instruments such as the EU Directives, though the signing of an international instrument may be binding for the national government. The laws are then normally implemented through a series of bylaws or regulatory measures taken by the administration and known in different jurisdictions as (ministerial/ presidential) decrees or statutory regulations. These are administrative measures of a regulatory nature that can only regulate issues covered by the mandate granted by the relevant law. These measures, when implemented by the public administration need to be further explained with a series of specific instructions as to their meaning and the ways in which they are to be applied. Such instructions are normally found in circulars. While circulars do not have a regulatory power or legal validity other than simple communication from and to the administration, in practice are the most important means of implementing a policy as these are the documents read by public servants in order to understand how the law is to be implemented.

In technical terms, such legal regulatory pyramid has an equivalent techno-regulatory pyramid. This comprises of international technical standards. At the top level of the techno-regulatory pyramid there are the international standards (for instance an ISO standard) which could then be adopted or specifically implemented in the national level by a national standardisation body (e.g. TSE). In some countries, especially in the context of e-government, there are specific processes as to how a standard could be introduced for the provision of e-government services or the interoperability between different e-government services. Normally the regulatory structure for the provision of web services standards and interoperability is as follows: a generic law provides for the issuing of ministerial decrees specifying the process which could be followed for the production of standards or the provision of interoperable services. Then a ministerial decree is issued specifying the process and an administrative agency or independent authority is specified to observe the standardisation and interoperability process. The standard may be proposed by a private or a public entity that may also be responsible for its maintenance but it is a body of experts that assesses the validity of the standard and the administrative body or independent authority that ensures such standard is adhered to in the production and provision of legal services. In addition, it is important that there is collaboration with all funding bodies that should refuse to fund any public projects unless they have the interoperability and standardisation criteria fulfilled.

Standardisation may also be mandated through topic specific regulation, e.g. medical or pharmaceutical related regulation, but as the systems will most probably make use of other e-government services (e.g. registries, social security etc), it is advisable that you go for horizontal interventions.

Possible Solutions

In the face of the previous observations with respect to the nature of the regulatory system, resolving the pharmacovigilance problem there are two ways to implement a solution:

Full legal structure implementation

In this case a full data protection law should be drafted and implemented. Such law should be compliant with the European Council DP Convention and follow the EU Data Protection Directive in order to make such law future proof. The fact that the DP Directive is about to be amended or that a DP EU Regulation is to be implemented should not be seen as a problem. Any future development in the EU level will certainly take into consideration the existing EU DP Directive and the EC DP Convention.

The law should also be complemented with:

- Administrative structures for the implementation and enforcement of the DP law
- Technical experts to provide specific guidance as to the technical specifications of systems making use of DP law especially in the context of PbD.
- Training of trainers with regards to the implementation of DP principles in the public administration.
- Training of DP officers in private organisations
- Introduction of DP seals depending on the level of protection granted to the consumer/ end-user

Implementation through a standardisation procedure

While the full legal implementation is the safest and most comprehensive way to tackle the problem, another way to proceed is through the use of technical standards that implement the PbD approach and follow the DP EC Convention. This approach still does not fully resolve the problem of having specific rules as to who and how should have access to the relevant data. However, it allows a much faster implementation of a solution, while potential inconsistencies may be avoided by closely following the DP EC and the EU DP principles, as well as their implementation in a way most relevant to the DP system Turkey would like to adopt.

The standardisation solution assumes the following steps:

- There is a ministerial decree or a law allowing a public body to oversee and implement a technical standard for the handling of either medical data in general or pharmaco-vigilance data in particular.

- Even in the case where there is no law specifically making reference to medical or pharmaceutical related data, there is great likelihood there is a law allowing the issuing of ministerial decrees for medical/ pharma issues in general. The existing ministerial decree was probably issued by virtue of such law.

- Based on this law there should be another ministerial decree specifically on the protection of personal data in the context of pharmacovigilance. Such decree should contain a standard or a procedure for the production of a standard that

would ensure the protection of personal data in this context. The standard should not refer only to high level DP principles but also to contain either the process for the production of compliance tables or compliance tables themselves. Compliance tables are tables containing technical specifications any vendor could follow in order to provide PbD systems. The compliance tables should be the result of policy decisions as to which DP principles are to be followed and how they are to be implemented in terms of more specific rules.

Early Conclusions

While the full legal solution is more consistent and future proof, it is also likely to take longer to be implemented and will require greater support from the government and the administration. The standardisation approach is quicker but will still require some form of specialisation of DP principles even if this is at the principles/ soft regulation and not the legislation level. This will then need to be translated into technical specifications and implemented through some form of administrative regulatory action. In both cases the solution will require some form of a study and collaboration between the civil service, jurists and a panel of experts or the TSE. Finally, the two solutions are not mutually exclusive: the standardisation procedure may be seen as a short term measure while the legislation is prepared in parallel or at a subsequent stage.

6. Private places: Protecting privacy against location providers

Authors: Maria Luisa Damiani and A Lawyer from the MODAP Privacy Observatory

Introduction

Computing and online services are increasingly being consumed through mobile devices, including smartphones and tablets. Indeed, more than half of the world's population now owns mobile phones, which are capable of running applications in ways that involve the collection, use and sharing of location data. Location based services (LBS) have become an integral part of users' experiences and an increasingly important market. They deliver users targeted, relevant and highly convenient information, such as up-to-the-minute traffic reports; the location of the nearest petrol stations, hospitals, or banks; as well as targeted advertisements and coupons for services located in a consumer's immediate range.

However, the significant advantages associated with LBS come at a price to users' privacy. While the location of a mobile device at any given moment may not be particularly sensitive, the historical trail of past locations can reveal much about a user's behaviour. People generally carry their mobile devices wherever they go, making it possible for LBS providers to pervasively track them. The use of location data without transparency and control is clearly troubling.

The Problem

Unfortunately, the architecture and inner workings of the LBS ecosystem remain opaque and largely unknown to users. Specifically, users often do not know that while they interact with and authorize a specific online or mobile application (App) to determine their location, such App turns to a location provider, typically Google, Apple or Skyhook, to obtain the localization service. Location providers allow mobile devices to determine location using a variety of methods, which include Global Positioning System (GPS), cellular triangulation, Wi-Fi and Internet Protocol (IP) based methods. Location providers compile extensive databases correlating Wi-Fi access points and cell phone antennas with their physical locations; mobile devices then query the database to obtain their current location. As repeat players interacting with numerous Apps, location providers are in a position to compile extensive records of users' location and movement.[11]

In 2010, W3C proposed a standard programming interface to enable the localization of App users[12]. The standard consists of a small set of functions which can easily be

[11] Soltani A., Testimony in front of the United States Senate, Judiciary Subcommittee on Privacy, Technology and the Law, Hearing on Protecting Mobile Privacy: Your Smartphones, Tablets, Cell Phones and Your Privacy, 10 May 2011, http://judiciary.senate.gov/pdf/11-5-10%20Soltani%20Testimony%20-%20Revised.pdf.

[12] Popescu A. (Ed.), W3C, Geolocation API Specification, Editor's Draft 10 February 2010, http://dev.w3.org/geo/api/spec-source.html.

embedded on an App. Notably, the standard requires explicit user consent for any instance of collection of location information by an App. The standard, however, fails to account for the fact users are not generally aware of the fact that their position is disclosed not only to Apps but also to location providers.[13]

Analysis

In its Opinion 13/2011 on Geolocation services on smart mobile devices, the Article 29 Working Party made clear that a user's location information constitutes "personal data" under the European Data Protection Directive, given its relation to an "identified or identifiable" individual.[14] Every mobile device has at least one unique identifier, the MAC address, which can be used to identify an individual user, even if her name is not known. Peter Hustinx, Europe's Data Protection Supervisor (EDPS) noted, "location data is certainly, in many instances, private data, and there then follows the obligations to inform users, and the opportunity to opt in or opt out." Under Article 4(1)(c) of the European Data Protection Directive,[15] the Directive applies even to location providers based outside of the EU, given that such entities "make use of equipment" (i.e., an end user's mobile device) situated in the EU.

Personal data may only be collected and processed based on a user's informed consent. The transfer of a user's location data to a third party who remains invisible to the user, violates the principles of transparency and user control prescribed by the European Data Protection Directive. Users do not have any reasonable means of obtaining information about, much less choosing among or consenting to the collection of their personal data by, location providers. Moreover, under existing architecture, users may convey to location providers personal data relating to third parties, namely individuals owning Wi-Fi equipment that registers on a user's mobile device. Needless to say, such third parties are not informed of or consent to the collection of their personal data.

To comply with the European Data Protection Directive, users must either be given clear and conspicuous notice of any prospective transfer of their information to a location provider, as well as the opportunity to opt in or out of the service. Alternatively, technological mechanisms must be put in place to prevent the possibility of location providers identifying or re-identifying a user based on her location data.

Possible solutions

---

[13] Doty, N.; Mulligan, D.K.; & Wilde, E., Privacy Issues of the W3C Geolocation API, UC Berkeley: School of Information Report 2010-038 (2010), http://escholarship.org/uc/item/0rp834wf.

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

[15] 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, 1995 O.J. (L 281) 31.

Like many other privacy and data protection problems, transfers of data to location providers need to be addressed through a combination of legal and technological mechanisms. *Legal tools* include the application of fair information principles, including transparency; user control; data minimization; and security. Users should be provided with clear, conspicuous and unavoidable notice concerning the transfer of their location information not only to Apps but also to location providers. In addition, users must be granted an opportunity to object to such data transfers, as well as periodic reminders of their right to withdraw their consent, and to choose among different location providers. Location providers should not collect personal data if not necessary to provide their service; refrain from logging location data in connection with unique device identifiers (UDID); and seek to de-identify data whenever possible. Location providers must collect data over encrypted channels and retain any personally identifiable information secure. (Indeed, these measures already appear, albeit implicitly, in the W3C standard, which provides: "The recipient of location information must not retransmit the location information without the user's express permission. Care should be taken when retransmitting and use of encryption is encouraged")[16].

Technological solutions can provide users even more robust privacy protections than legal rules. We suggest an approach based on caching information that is useful to determine a user's position in proximity of any self-identified "private locations" which she has already visited. We call this approach a "privacy-aware geolocation strategy"[17]. A privacy-aware geolocation strategy is based on the realization that the amount of information currently conveyed by users to location providers exceeds that necessary to determine their exact position. In order to minimize the interaction between a user and a location provider, specific locations will be delineated "private" by users, and thereafter not reported to location providers. Whether a place is private or not depends exclusively on an individual's perception and not on physical characteristics (*e.g.*, indoor or outdoor). Hence, the approach is to extract from the concept of place that of a "private place". A private place conceptualizes the intuition that there are some locations (*e.g.*, home; psychologist's office; boyfriend's street) which belong to the personal sphere. Users' positions inside private places should not be disclosed to location providers.

Conclusion

In this opinion report we advocate implementing a combination of legal and technological tools to enhance users' privacy and increase transparency and control with respect to the collection of location data by third party location providers. We propose an innovative approach, based on empowering users to "mask" specific areas they deem "private locations" from scrutiny by location providers.

[16] Popescu, supra note 12.

[17] Damiani M. L., Third party geolocation services in LBS: privacy requirements and research issues, Transactions on Data Privacy (2011) 1–18.

7. Big data mining, fairness and privacy: *A vision statement towards an interdisciplinary roadmap of research*

Authors:
Dino Pedreschi, Knowledge Discovery and Data Mining Lab, University of Pisa
Toon Calders, Information Systems Group, Eindhoven University of Technology
Bart Custers, eLaw - Centre for Law in the Information Society, Leiden University
Josep Domingo-Ferrer, UNESCO Chair in Data Privacy, Universitat Rovira i Virgili
Giusella Finocchiaro, Department of Law, University of Bologna
Fosca Giannotti, Knowledge Discovery and Data Mining Lab, ISTI-CNR, Pisa
Morag Goodwin, Tilburg Institute for Law, Technology, and Society, Tilburg University
Mireille Hildebrandt, Erasmus School of Law, Rotterdam and Centre for Law Science Technology and Society Studies, Vrije Universiteit Brussel
Stan Matwin, School of Information Technology and Engineering, University of Ottawa
Yucel Saygin, Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul
Bart Schermer, eLaw - Centre for Law in the Information Society, Leiden University
Tal Zarsky, Faculty of Law, Haifa University

We live in times of unprecedented opportunities of sensing, storing and analyzing micro-data on human activities at extreme detail and resolution level, at society scale. Wireless networks and mobile devices record the traces of our movements. Search engines record the logs of our queries for finding information on the web. Automated payment systems record the tracks of our purchases. Social networking services record our connections to friends, colleagues, collaborators.

Ultimately, these **big data of human activity** are at the heart of the very idea of a knowledge society: a society where decisions – small or big, by businesses or policy makers or ordinary citizens – can be informed by reliable knowledge, distilled from the ubiquitous digital traces generated as a side effect of our living. Increasingly sophisticated data analysis and data mining techniques support knowledge discovery from human activity data, enabling the extraction of models, patterns, profiles, simulation, what-if scenarios, and rules of human and social behavior – a steady supply of knowledge which is needed to support a knowledge-based society. The *Data Deluge* special report in "The Economist" in February 2010 witnesses exactly how this way of thinking is now permeating the entire society, not just scientists.

The capability to collect and analyze massive amounts of data has already transformed fields such as biology and physics, and now the human activity data cause the emergence of a data-driven "computational social science": the analysis of our digital traces can create new comprehensive pictures of individual and group behavior, with the potential to transform our understanding of our lives, organizations, and societies.

The paradigm shift towards human knowledge discovery comes, therefore, with unprecedented opportunities and risks: *we the people are at the same time the donors of the data that fuel knowledge discovery and the beneficiaries – or the targets – of the resulting knowledge services*. The paradoxical situation we are facing today, though, is that we are fully running the risks without fully grasping the opportunities of big data: on

the one hand, we feel that our private space is vanishing in the digital, online world, and that our personal data might be used without feedback and control; on the other hand, the very same data are seized in the databases of global corporations, which use privacy as a reason (or excuse?) for not sharing it with science and society at large.

In the CNN show *Fast Future Forward*, the anchor-woman asked the panel of futurists: "Which major challenge are we going to have to deal with 10 years out?" The answer was: "We'll have to completely reverse our orientation to privacy. The reality is that we don't have privacy anymore: you use your cell phone, you drive your car, you go on-line, and it's gone." Although the debate on the vanishing privacy is going on for years, it is striking that now it is posed as a major problem in a popular prime-time show on the future trends of society and technology. The message is clear: privacy, once a human right, is becoming a chimera in the digital era.

In the other extreme, knowledge discovery and data science run the risk of becoming the exclusive and secret domain of private companies – Internet corporations such as Google, Facebook, Yahoo, or big telecom operators – and government agencies, e.g. national security. For these data custodians, privacy is a very good excuse to protect their interests and not share the data, while users are not really aware how the data they generated are used. Alternatively, there might emerge a cast of privileged academic or industry researchers who are granted access over private big data from which they produce results that cannot be assessed or replicated, because the data they are based on cannot be shared with the scientific community. Neither scenario will serve the long-term public interest of accumulating, verifying, and disseminating knowledge.

Should we really give up and surrender to a wild digital far west, where people are exposed to risks, without protection, transparency, and trust, while society as a whole and science get little reward with respect to the opportunities offered by the secluded big data? We believe that another knowledge technology is possible, a fair knowledge discovery framework can be devised, where opportunities are preserved and risks are kept under control. **A technology to set big data free for knowledge discovery, while protecting people from privacy intrusion and unfair discrimination**.

In order to fully understand the risks, we should consider that the knowledge life-cycle has two distinct, yet intertwined, phases: *knowledge discovery and knowledge deployment*. In the first step, knowledge is extracted from the data; in the second step, the discovered knowledge is used in support of decision making; the two steps may repeat over and over again, either in off-line or in real-time mode. For instance, knowledge discovery from patients' health records may produce a model which predicts the insurgence of a disease given a patient's demographics, conditions and clinical history; knowledge deployment may consist in the design of a focused prevention campaign for the predicted disease, based on profiles highlighted by the discovered model. Hence, people are both the data providers and the subjects of profiling. In our vision, the risks in each of the two steps in the knowledge life-cycle are:

- *Privacy violation*: during knowledge discovery, the risk is unintentional or deliberate intrusion into the personal data of the data subjects, namely, of the (possibly unaware) people whose data are being collected, analyzed and mined;

- *Discrimination*: during knowledge deployment, the risk is the unfair use of the discovered knowledge in making discriminatory decisions about the (possibly unaware) people who are classified, or profiled.

Continuing the example, individual patient records are needed to build a prediction model for the disease, but everyone's right to privacy means that his/her health conditions shall not be revealed to anybody without his/her specific control and consent. Moreover, once the disease prediction model has been created, it might also be used to profile the applicant of a health insurance or a mortgage, possibly without any transparency and control. It is also clear, from the example, how the two issues of profiling and privacy are strongly intertwined: the knowledge of a health risk profile may lead both to discrimination and to privacy violation, for the very simple fact that it may tell something intimate about a person, who might be even unaware of it.

Privacy intrusion and discrimination prevent the acceptance of human knowledge discovery: if not adequately countered, they can undermine the idea of a fair and democratic knowledge society. The key observation is that they have to be countered together: focusing on one, but ignoring the other, does not suffice. Guaranteeing data privacy while discovering discriminatory profiles for social sorting is not so reassuring: it is just a polite manner to do something very nasty. So is mining knowledge for public health and social utility, if, as a side effect, the personal sensitive information that feeds the discovery process is disclosed or used for purposes other than those for which it has been collected, putting people in danger. On the contrary, protecting data privacy and fighting discrimination help each other: methods for data privacy are needed to make the very sensitive personal information available for the discovery of discrimination. If there is a chance to create a trustworthy technology for knowledge discovery and deployment, it is with a holistic approach, not attempted so far, which faces privacy and discrimination as two sides of the same coin, leveraging on inter-disciplinarity across IT and law. The result of this collaboration should enhance trust and social acceptance, not on the basis of individual ignorance of the risks of sharing one's data, but on a reliable form of risk measurement. By building tools that provide feedback and calculated transparency about the risk of being identified and/or discriminated, the idea of consent and opt-in may become meaningful once again.

In summary, a research challenge for the information society is the definition of a theoretical, methodological and operational framework for fair knowledge discovery in support of the knowledge society, where fairness refers to privacy-preserving knowledge discovery and discrimination-aware knowledge deployment. The framework should stem from its legal and IT foundations, articulating data science, analytics, knowledge representation, ontologies, disclosure control, law and jurisprudence of data privacy and discrimination, and quantitative theories thereof. We need novel, disruptive technologies for the construction of human knowledge discovery systems that, **by design**, offer native techno-juridical safeguards of data protection and against discrimination. We also need a new generation of tools to support legal protection and the fight against privacy violation and discrimination, powered by data mining, data analytics, data security, and advanced data management techniques.

The general objective should be the reformulation of the foundations of data mining in

such a way that privacy protection and discrimination prevention are embedded the foundations themselves, dealing with every moment in the data-knowledge life-cycle: from (off-line and on-line) data capture, to data mining and analytics, up to the deployment of the extracted models. We know that technologies are neither good nor bad in principle, but they never come out as neutral. Privacy protection and discrimination prevention have to be included in the basic theory supporting the construction of technologies. Finally, the notions of privacy, anonymity and discrimination are the object of laws and regulations and they are in continuous development. This implies that the technologies for data mining and its deployment must be flexible enough to embody rules and definitions that may change over time and adapt in different contexts.

The debate around data mining, fairness, privacy, and the knowledge society is going to become central not only in scientific research, but also in the policy agenda at national and supra-national level. We conclude our discussion by proposing a list of the top ten research questions, as a contribution to set the research roadmap around fair knowledge discovery and data mining.

1. How to define fairness: how to actually measure if privacy is violated, identity is disclosed, discrimination took place?

2. How to set data free: how to make human activity data available for knowledge discovery, in specific contexts, while ensuring that freed data have a reliable measure of fairness?

3. How to set knowledge free, while ensuring that mined knowledge has been constructed without an unfair bias? How to guarantee that a model does not discriminate and does not compromise privacy?

4. How to adapt fairness in different contexts, which raise different legitimate expectations with regard to privacy and non-discrimination? To what extent should discrimination on the basis of a higher risk be defined as discriminatory? Is this a legal issue or an ethical issue?

5. How to make the data mining process parametric w.r.t a set of constraints specifying the privacy and anti-discrimination rules that should be embedded in freed data and mined knowledge? How to take into account in the process the analytical questions that will emerge after having mined the data?

6. How to prove that a given software is fair, if it is claimed to be fair? How can the relevant authorities check whether e.g. a privacy policy or a non-discrimination policy is complied with?

7. What incentive does the industry have to opt for fair data mining technologies? What incentive do individual consumers have to opt for service providers that employ such technologies?

8. How to provide transparency about services that employ profiling and are potentially discriminatory? Which effective remedies do citizens have if they suspect unfair discrimination on the basis of data mining technologies (cf. art. 12 Directive 95/46/EC and the Anti-Discrimination Directives 2000/43/EC

2000/78/EC, 2004/113/EC, 2006/54/EC)?

9. Which incentives could be provided by the European legislator to employ fair data mining technologies (both on the side of the industry and on the side of individual consumers)? E.g., compulsory certification, new legal obligations, technical auditability schemes, new individual rights, new data protection competences?

10. How to specify and implement fairness on-line, so as to guarantee privacy and discrimination freedom at the very moment of data acquisition or capture?

8. The rise of identifiability: The downfall of personal data?

Authors: Colette Cuijpers Yucel Saygin

Abstract accepted to CPDP 2012 (Computers, Privacy and Data Protection) Conference. Full paper was provided in the appendix

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*" (WP 136: 21). 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 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 since plenty of background information exists 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. This research clearly demonstrates how geolocation information can fairly easily be combined with other open source information, turning it into identifiable and thus personal information. This characteristic of geolocation information is acknowledged by the Article 29 Working Party in WP 185. In this article we will assess whether the sketched development should bear influence upon the concept of personal data. The assessment will not only be given from a legal perspective, reconsidering the definition and interpretation of this concept, but also from a technical perspective as *"putting in place the appropriate state-of-the-art technical and organizational measures to protect the data against identification may make the difference to consider that the persons are not identifiable"* (WP 185). The technical assessment will focus on technical measures to prevent identifiability in relation to geolocation information.

9. Show me who your friends are and I'll tell you who you are: Security flaws in facebook's privacy settings

Authors: Andrea Zanda, Yucel Saygın, and a Lawyer from the Privacy Observatory

Introduction

One of the most significant developments in the online environment over the past few years has been the meteoric rise of social networking services. Facebook, first launched in 2004, had 100 million users in late 2008 and saw that number increase eightfold in a matter of just three years. Twitter, created in 2006, had 100 million registered users in 2010, and was recently rated by Alexa as the 9th most popular web site, five places ahead of online giant Amazon.

Online users post massive amounts of personal information online, including photos and videos; friends' lists; location information; political, ideological and artistic tastes; social relationships and sexual interests. At the same time, social networking services boast about respecting users' preferences by maintaining granular privacy controls, thereby allowing users to control just how much information they would like to share.

In this report we demonstrate that despite Facebook's promises in its Privacy Settings, users' friends' lists can be discerned by sophisticated third parties even where users choose to keep them private.

The Problem

Facebook users can choose whether their friends' list is shared with the "public" (meaning the open Internet); Facebook friends only; or remains completely private ("only me"). In addition, users can organize their friends' lists into groups and customize more finely grained privacy preferences with respect to the visibility of those lists by other users or the entire Internet. However, we discovered that through the Facebook developers' Graph API, we can verify whether or not two random users are each other's Facebook friends. We demonstrate this in the figure below. We tested our claims with two Facebook accounts belonging to Tali Fink and Omer Tene, user ids 100001167521942 and 739554292 respectively. The method call "friends.areFriends?"[18] utilizes the user ids as parameters and queries the Facebook graph. The result of the query for the specified user ids appears in the bottom right rectangle: "are_friends: true". This means that the two users with the given ids are indeed Facebook friends. This

[18] http://developers.facebook.com/docs/reference/rest/friends.areFriends/

functionality enables attackers with some background information to reconstruct users' private friends' lists and sparse out their social graph.

Analysis

It is troubling to find that Facebook users' privacy settings can easily be circumvented to reveal their social graph. To be sure, random crossing of Facebook user names in the hope of finding pairs of friends would be an exercise in futility, as the number of possible combinations is endless. Yet far more targeted attacks are easy to envisage, for example utilizing the public friends list of user A, a Facebook friend of user B who keeps her own friends list private, to make educated guesses about the identity of friends of user B. Or, to take another example, identifying lists of political activists based on cross friendships.

Conveying to users a false sense of security with respect to their privacy choices contradicts the principles of notice and choice, which govern online privacy in both the United States and the European Union. Moreover, we view a user's friends' list as potentially sensitive information. As the saying in the title posits, much can be learned about an individual by analyzing her list of friends. For example, computer science students at MIT demonstrated that by looking at a user's Facebook friends they could predict whether the user was homosexual.[19] And the Google Buzz fiasco, which cost Google $8.5 million in a class action settlement, involved Gmail users being outraged upon finding that their contacts were "outed" by Google – converted to recommended friends' lists on a newly launched social network.[20]

[19] Carolyn Johnson, Project 'Gaydar': At MIT, an experiment identifies which students are gay, raising new questions about online privacy, Boston Globe, September 20, 2009, http://www.boston.com/bostonglobe/ideas/articles/2009/09/20/project_gaydar_an_mit_experiment_raises_new_questions_about_online_privacy.

[20] Amir Efrati, Google Settles Privacy Lawsuit for $8.5 Million, Wall Street Journal, Sept. 3, 2010, http://online.wsj.com/article/SB10001424052748703946504575470510382073060.html.

Possible solution

We suggest that Facebook either fixes the apparent security fault identified above or clearly explains to users that through individual analysis, even private friends' lists can be discerned by interested third parties.

10. Open data and Privacy: two worlds collide?

Authors: Manolis Terrovitis and Spiros Athanasiou

Introduction

Human activities are becoming increasingly digitized. Business and financial activities reside completely inside information systems, the distribution of goods is traced by RFIDs, moving vehicles are monitored by GPS tracking devices, energy production and consumption is measured in real time by digitized meters, even social interaction has a vast digital aspect in the online social networks. A huge volume of data is collected by various organizations and companies in order to monitor their operations and it is used as a resource for providing knowledge services. This process has been recognized as the emergence of Big Data in the ICT technologies, which is expected to create a new business sector in the next few years.

The digitization of human activities has not only precipitated the challenge of Big Data but it has sparked the Open Data trend around the world. Recognizing data, and especially Big Data, as means to transparency and also as an important resource, NGOs, public administrations and the EU have promoted open access to data, establishing Open Data as a default practice in several fields.

The challenges around Big and Open Data are numerous. There are technical challenges including the efficient processing, storing and re-using of huge data volumes and also the creation of publication models that can facilitate the meaningful dissemination of Big Open Data. There are also legal, financial and social challenges, which focus on how to create incentives for all stakeholders of the data production circle to share their data, on how to create new business models for profiting from data, and how to guarantee the rights of each stakeholder in the data processing. Addressing these challenges is expected to have a positive impact on productivity, transparency and social participation.
Big Open Data do not come without any strings attached; the detailed monitoring of personal and business activities poses significant threats to personal privacy and also to corporate privacy. To promote Open Data and to remove restrictions in processing Big Data in general, we need tools that will allow guarantee the protection of privacy while preserving the data quality needed for scientific, business and other uses.

The Open Data Landscape

Open data, i.e. data that can be freely used, extended and redistributed, is an established scientific practice, necessary for scientific advancement. In the past years however, open knowledge practices have spread beyond the scientific realm and into the mainstream political and technological agenda. A new movement has been formed, the *open-data movement*. The reason for this development is the realization of the tangible benefits open data provide.
On a political level, open data is a facilitator for transparency, accountability and participation, i.e. core democratic values. Access to open data can empower citizens to

collectively audit and participate in policy making. On the public administration level, open data can provide cost reductions through increased technical and semantic interoperability, immediate access to information for better decision making, and ultimately more efficient services. On an economical level, open data can boost growth, enabling the private sector to provide products and services of lower cost and increased quality, foster innovation, deliver value added services, and establish an emerging economic activity, identified as the *data economy*. The Vickery report estimates the benefits of extended PSI (Public Sector Information) reuse for the EU27 economy at 140b€/year (1.7% GDP 2008). These financial benefits will be materialized through the establishment of a data economy, led by SMEs providing added value services by repurposing and extending open data.

Consequently, a number of policies have been implemented in the EU to promote open data. Directive 2003/98/EC on the reuse of Public Sector Information (PSI), and its recent revision proposal to further open up the market for services based on PSI, establish a common binding legislative framework in EU member states for open data & information. Directive 2007/2/EC (INSPIRE) establishes a common legislative, organizational and technical framework to promote interoperable Spatial Data Infrastructures (SDIs) across member states and common sharing of environmental data. Further, the Digital Agenda for Europe and the EU 2020 Strategy imprint a pan-European strategic goal on open data as a facilitator for economical growth, innovation and better governance. On a world-wide level, open data are promoted through the Open Government Partnership, while several international organizations provide high-value open data (e.g. UN, World Bank).

Early open data initiatives include the UK and USA open data portals (data.gov.uk and data.gov respectively), which provided guidance, publishing guidelines and reusable software components. Currently, practically all EU member states have official and/or volunteer driven open data portals. A pan-European open data catalogue which will harvest and provide a single point of access to all of EU's open data is already in operation (publicdata.eu). The quantity and quality of published open data is constantly growing: more than 30.000 open data sets are available in EU alone and include data from multiple domains (e.g. spending, environment, geospatial). The emphasis is place on *high-value data*, where *value* is expressed both as *financial* value and *significance* for good governance.

Privacy as an afterthought

Public concern about user privacy on the internet has significantly increased in the last few years. Awareness of the threats against sensitive personal data has increased in public administrations, researchers and companies, but no widely accepted framework for privacy protection has been established.

Threats against user privacy in the digital world and especially on the web, are a cause of growing public concern. Several inherent characteristics of the web, like replicability, searchability and persistence have disrupted users' ability to regulate their audience.

Information on the web easily reaches audiences who were not supposed to receive it, without losing any accuracy. Moreover, information from various sources can be linked and associated with a certain individual, revealing as a whole, significantly more than the user intended.

The dangers against user privacy are not easily identified by the public, but also by companies, public administrations and even researchers. Existing policies and solutions are at their infancy and they are only partially applied. Privacy enhancing technologies and privacy-by-design systems are few and they cannot still cope with the variety of dangers that emerge as more and more data become available. Legal frameworks lag even more behind the technical solutions. Moreover, as information on the web transcends international boundaries, enforcing any local policy for privacy protection is infeasible.

Research in privacy preserving technologies has been very active in the last few years and significant results have emerged. Secure protocols for exchanging data over social networks, private information retrieval, privacy-by-design principles in software engineering and anonymization are just a few of the promising ones. Anonymization techniques are especially promising, since instead of creating a secure workflow for data sharing, they alter the data themselves to remove sensitive information or to prohibit its linking to individual even with the help for external knowledge sources. Still, these solutions cover only few of the known dangers and industry is only beginning to adopt them.

Privacy issues are often considered as a constraining factor for the adoption of open data, and arguments against openness are often based on the protection of sensitive personal data. Still, even a small survey of the reported privacy breach incidents, shows that the greatest dangers usually stem from social networks, from information that users have unintentionally disseminated to undesired audiences and from security breaches in companies and organizations. Open data are usually offered by public administrations and in most cases they have been anonymized, even if trivially. Without dismissing the dangers to privacy it is fair to say that until now the dangers posed by providing open data, are far overshadowed by the benefits.
Still the dangers should not be ignored: as open data increase the dangers will become even more significant. If no measure is taken, the success and the widespread adoption of open data might backlash. It is not impossible to balance between public interest and privacy protection. In several cases there are promising technical solutions (e.g. anonymization), but in others it is mostly a matter of perception, policy, and legal framework (e.g. Diavgeia, the Clarity Programme, et.diavgeia.gov.gr). To reach to such a balance as soon as possible we believe that public administrations together with all stakeholders in the open data movement, should actively take initiative to bolster research efforts and public awareness in matters of privacy protection.

A real world example

In the following we examine a real-world example, where the availability of open data in conjunction with lacking technical decisions introduced a significant privacy risk. The provided example took place in Greece, during October 2010 and has remained unpublished until now. The relevant authorities have been notified for our findings within hours and measures to address the risks were implemented within weeks.

Since 1982, vehicles can enter the city of Athens, Greece depending on the last digit of their license plate number and the current calendar day. Specifically, vehicles with a license plate number ending on an *even* number can enter the center of Athens during the *even*-numbered calendar days, and vice-versa for odd-numbers. For example, the vehicle with a license plate IMB121**2** (which ends with a '2', thus an even number) can enter in the 2nd, 4th, day of the calendar month. The system is called "odd/even" and aims to reduce traffic congestion and pollution. It has been extremely successful because of its simple enforcement; a traffic warden can check the last digit of a moving vehicle within seconds and take action if necessary.

In 2012, the Ministry of Environment, Energy, and Climate Change, decided to alter this scheme. In particular, the goal was to grant low-polluting cars (i.e. vehicles with Euro 4/5 engines and CO2 emissions < 140g/km) the right to enter the center of Athens *at all days*. Such a move would serve as a motivation for citizens to invest in energy efficient, clean vehicles. However, the scheme introduced challenges regarding its enforcement. Traffic wardens lost the ability to perform a split second decision based on the license plate number alone; they also required information regarding the engine type and emissions of the vehicle. However, there was no available technical infrastructure (e.g. application for mobile terminals), nor the required funds to print and issue stickers for vehicles (costs would rise to 40M euros).

To address these problems, the Ministry of Environment designed the following solution. All vehicle owners of the city of Athens (appr. 2M) would be requested to enter a web site, check if their vehicle is considered as a *clean car* according to the new scheme, and if yes, download and print *themselves* a small sticker. Of course this was a no-solution since everyone could print within 5min a counterfeit sticker. In hindsight, vehicles in the city of Athens carrying the stickers now are almost inexistent. However, the actual problem lies within the technical implementation of the proposed solution.

A simple web application (based on WordPress) was developed and advertised across digital and print media in Greece. Citizens were required to enter their personal VAT and vehicle license plate number. A G2G web service developed by the General Secretariat of Information Systems was invoked under the scenes, and provided the application with detailed information regarding the *vehicle and its owner*. Depending on the engine type

and emissions, a response was provided to the user which included (among other personal data) the owner's *VAT* number, and the vehicle's *VIN* number.

This application had many technical flaws and omissions (e.g. no user login, no use of https), of which the most important was the complete *lack of sessions*. In simple terms, the url informing citizens regarding the result of their query (i.e. *is my car eligible*) was available for everyone to access in a *static url*. Consequently, even novice programmers could easily write a simple scrapper that would constantly invoke this url and store its contents. Considering that almost 2M vehicle owners were requested to use the web application, one could compile within weeks a database with the *associated VAT and VIN numbers for almost 2M citizens*.

This produced database is clearly an exploit resulting from lacking technical skills; similar privacy risks have been documented in the past. However, these risks can now be *leveraged at frightening scales* by combining them with open data. In this particular example, the produced database could be collated with two open data sources: (a) an open web service developed by the General Secretariat of Information Systems that provided information (including residential address) for an individual based on their VAT number, and (b) multiple open web services providing specific details on a vehicle based on its VIN number (e.g. vindecoder.net). Consequently, one could automatically combine all the above data to discover *expensive* vehicles (based on their VIN number) and the *address* of their owner (based on the VAT number).
As demonstrated by this real world example, the introduction of open data into the mainstream ICT landscape can have significant and unforeseen consequences, far beyond their intended use.

Conclusion

The Open Data practice has already exhibited tangible economic, societal and political benefits, by promoting transparency, accountability and participation and also by removing obstacles to valuable knowledge. The warm support of the public, of the EU and of many public organizations is a result of the immediate gains that stem from sharing data openly. In the next years we expect that more and more data will be available as Open Data and that their impact on the economy will be paramount, surpassing that of Big Data#.

Privacy related issues comprise a growing concern that relates with Open Data. Still, we believe that in the unregulated landscape of web, where no established practices for protecting user and corporate privacy, Open Data constitute a low risk factor. This does not imply that it can be ignored. As Open Data are increasingly becoming available the privacy risks will grow. Moreover, since the publishers of Open Data are usually public organizations, it is expected that they will set up the standard for privacy protection.

Reconciling Open Data and Privacy protection is not an impossible task; already several privacy enhancing technologies exist, with anonymization being the most promising, to allow sharing data without revealing sensitive information. Still, both technical and legal tools for privacy protection lag behind the advances in information retrieval and processing. To establish a balance between data exploitation and privacy protection significant initiatives must be undertaken in research and in practice to create the suitable tools and to increase public awareness.

BNA

Privacy and Security Law Report®

Privacy and Data Protection in Turkey: (Inching) Towards a European Framework

BY OMER TENE AND YUCEL SAYGIN

Introduction

With a population of 73 million, second only to Germany when compared to the 27 European Union Member States, and a gross domestic product growth rate of 8 percent, Turkey is a force to be reckoned with on the outskirts of Europe. Turkey harbors the largest city in Europe (Istanbul, with an estimated population of 15 million); has 63 million mobile telephone subscribers; and is the fifth largest country in the world in terms of the number of Facebook users (after the United States, Indonesia, India, and the United Kingdom) with 30 million.

Yet despite many years of political and public debate, Turkey remains without a comprehensive legal framework for privacy and data protection. Besides constituting an obstacle in Turkey's quest to become a member of the European Union, the lack of a modern data protection law limits business opportunities and potential collaboration between Turkish and foreign businesses and law enforcement entities and compromises the fundamental rights of Turkish individuals.

Omer Tene is managing director, Tene & Associates, and Associate Professor, Israeli College of Management School of Law. Yucel Saygin is a professor of engineering and natural sciences at Sabanci University, Istanbul. Both are participants in MODAP (Mobility, Data Mining, and Privacy), a Coordination Action project funded by the EU FET OPEN.

Constitutional Amendments

In March 2010, the Turkish government submitted to Parliament a package of constitutional amendments, including an amendment to the privacy protection provision, Article 20, adding a section on data protection.[1] The constitutional amendments were authorized by the unicameral Turkish Grand National Assembly in May 2010, yet failed to garner the support of a two-thirds majority required to pass directly, gaining 336 of the 500 members of Parliament, short of the 367 threshold. They were therefore submitted to a nationwide referendum in September 2010, after withstanding constitutional challenge in the Turkish Constitutional Court, where they passed with sweeping support of 58 percent of the vote.

Pursuant to the constitutional amendment, Article 20 now states that ''everyone shall have the right to the protection of their personal data.''[2] It provides that personal data may only be collected, stored, and used pursuant to a legal obligation or with the data subject's consent. It grants individuals the right to be informed about any processing of their personal data, as well as the right to access, rectify or delete their personal data. It calls for legislation setting forth the principles and procedures to implement the constitutional mandate. Yet even before such legislation is put in place, Turkish individuals benefit from constitutional protection *vis-à-vis* action by public sector bodies. Indeed, in its latest progress report, the European Commission states that

[1] The text added to Article 20 of the Turkish Constitution provides: ''Everyone has the right to demand the protection of his personal data. This right comprises the right to be informed about the personal data concerning himself, access to such data, right to request correction or deletion of them as well as the right to be informed if such data is [sic] used in accordance with the purposes for which it was collected. Personal data can be processed only in cases regulated in a law and upon express consent of the subject individual. Principles and procedures regarding the protection of personal data shall be regulated by a law.''

[2] Communications secrecy continues to be protected by Article 22 of the Constitution.

''with regard to fundamental rights, progress has been made. Constitutional amendments bring important changes in the area of data protection (. . .).''[3]

Finally, Turkey is currently in the process of introducing a new Constitution to replace the existing document, which dates from 1982, a relic of the 1980 *coup d'état*.[4] Given that privacy and data protection are not politically controversial, we hope that the new Constitution will echo the provisions of the recently amended Article 20.

Omnibus Legislation

Omnibus data protection legislation has been contemplated in Turkey for almost a decade, but has been slow in coming. Turkey is a signatory to the Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (''Convention 108''), yet has not ratified the Convention. It is has also been a member of the Organization for Economic Co-operation and Development (OECD) since 1961. As part of Turkey's 2005 Accession Partnership with the European Union, Turkey was required to ''adopt a law on protection of personal data'' and ''establish an independent supervisory authority.''[5]

According to the European Commission's 2009 progress report, Turkey is yet to take the required measures to implement this data protection agenda. The Commission stated: ''With regard to respect for private and family life and, in particular, the right to protection of personal data, Turkey needs to align its legislation with the data protection *acquis*, in particular Directive 95/46/EC, and, in that context, to set up a fully independent data protection supervisory authority. Turkey also needs to ratify both [Convention 108] and the additional protocol to it on supervisory authorities and transborder data flow (CETS No 181). Wiretapping has reportedly been used extensively and records of it published in the press.''[6] In its 2010 progress report, the Commission emphasized that an ''effective personal data protection regime is crucial for efficient international judicial cooperation.''

In 2003, the Turkish Ministry of Justice tabled a Draft Law on the Protection of Personal Data (the ''Draft Law''); yet the Draft Law has since been lost in parliamentary committee hearings, while Turkey remains without proper legislation. The Draft Law is predicated on the European Data Protection Directive (95/46/EC). It applies ''to natural and legal persons whose personal data are processed'' and defines personal data as ''any information relating to an identified or identifiable natural or legal person.'' Hence, unlike most European data protection laws, the Draft Law would protect both individuals and legal entities. Similar to European data protection legislation, it applies to both automated and manual data processing. The Draft Law sets forth principles for the processing of personal data, including purpose specification, proportionality, accuracy, and retention limitation. It lists legitimate bases for processing personal data, including consent and compliance with a legal obligation.

Article 7 of the Draft Law, which applies to sensitive data, permits processing ''[i]f the personal data are required to be processed for purposes of preventive medicine, medical diagnosis, medical treatment, health care or execution of health services by: (1) Health facilities, (2) Insurance companies, (3) Social security institutions, (4) Employers with the liability of establishing [a] health unit at the work place, (5) Schools and universities; in accordance with the relevant laws, under the supervision of health care personnel (. . .).'' We believe that this provision, which is based on Article 8(3) of the European Data Protection Directive, is overly broad. It exposes highly sensitive data not only to health care providers but also to insurance companies, social security institutions, employers and schools; indeed, it is difficult to foresee *any* use of medical data that is not authorized by Article 7. As such, it lacks adequate safeguards to protect the privacy of patients and their families. Hence, even if it is pushed through the legislative process, the Draft Law remains deficient.

Other provisions in the Draft Law appear internally inconsistent. For example, under Article 5(2) of the Draft Law, ''[p]ersonal data can be re-processed for scientific or statistical purposes or for a longer duration than stated (. . .), provided that there [are] adequate safeguards in the relevant legislation regarding the purpose of re-processing (. . .).'' Thus, this provision allows processing of data for secondary purposes, including scientific or statistical purposes, without expressly requiring that such data be anonymized. At the same time, Article 10 of the Draft Law provides that ''[f]or research, planning and statistical reasons which aim at public benefit, the personal data can be processed provided that they are made anonymous.''

Additional provisions of the Draft Law provide individuals with rights of transparency, access and rectification (Articles 11–13) and impose data security obligations (Article 15), as well as restrictions on cross-border data transfers based on the principle of adequacy (Article 14). Certain provisions of the Draft Law already appear outdated, such as those setting up a registry of databases (Articles 16–19). Even in Europe, the bedrock of the database registration (notification) regime, regulators have come to accept that in a globalized economy with ubiquitous data flows, such bureaucratic exercises yield few tangible benefits. The Draft Law further sets up a data protection authority as well as an independent data protection board (Articles 26–33), although the relation between the two is not entirely clarified by the legislative text. Finally, the Draft Law sets forth penal, administrative and civil causes of action for willful or negligent unlawful processing of data.

---

[3] Commission Staff Working Document, Turkey 2010 Progress Report, accompanying the Communication from the Commission to the European Parliament and the Council, Enlargement Strategy and Main Challenges 2010-2011, 9 November 2010, http://ec.europa.eu/enlargement/pdf/key_documents/2010/package/tr_rapport_2010_en.pdf.

[4] Ece Toksabay, the Turkish prime minister, wants new Constitution by the first half of 2012, Al Arabiya News, Sept. 29, 2011, available at: http://www.alarabiya.net/articles/2011/09/29/169270.html.

[5] Council Decision 2008/157/EC of 18 February 2008 on the principles, priorities and conditions contained in the Accession Partnership with the Republic of Turkey and repealing Decision 2006/35/EC, available at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:051:0004:01:EN:HTML.

[6] Commission Staff Working Document, Turkey 2009 Progress Report, accompanying the Communication from the Commission to the European Parliament and the Council, Enlargement Strategy and Main Challenges 2009-2010, 14 October 2009, http://ec.europa.eu/enlargement/pdf/key_documents/2009/tr_rapport_2009_en.pdf.

   ISSN 1538-3423

Almost a decade has passed since the introduction of the Draft Law—a very long time in terms of technological progress and business development. Innovations and breakthroughs, particularly in information and communications technologies, have created new business models and tools affecting individuals' lives and impacting the functioning of virtually every business process and government activity. This means that the Draft Law is in danger of becoming obsolete even before its enactment. Indeed, the EU Data Protection Directive, which serves as its main model, is subject to comprehensive review by the European Commission this year. In devising its data protection legislative strategy, Turkey should take account of these transformations to avoid adopting an outdated scheme.

Sector-Specific Regulation

With omnibus data protection legislation stalled and the establishment of a data protection authority delayed, Turkey has begun developing sector-specific regulatory schemes. First and foremost is the framework for the protection of privacy and personal data in electronic communications. In 2000, Turkey established the Telecommunications Council as the institution responsible for data protection in the telecom sector. The Council was responsible for the enactment in 2004 of the By-Law on Personal Data Processing and Protection of Privacy in the Telecommunications Sector (the ''Regulation'') as well as the 2008 By-Law on Security of Electronic Communications. The Regulation is based on the principles of the European e-Privacy Directive (2002/58/EC), such as communication confidentiality and security, the handling of traffic and location data, call number identification, public subscriber directories, and the prevention of spam.

In the past few years, the Council, now named the Information and Communications Technology Authority (''ICTA''), has enforced the Regulation forcefully, imposing steep penalties and making its presence felt in the telecom sector. In one case, *Taraf*, a Turkish newspaper, disclosed that personal data of subscribers of a telecom operator could be accessed by third parties. After a thorough investigation, ICTA issued the operator a fine in an amount of TL 1.25 million ($800,000). Another case involved illicit access by a former Turkish soccer star, Ridvan Dilmen, to call traffic records of his ex-girlfriend. An investigation launched by the ICTA and carried out by the police ultimately led to the imposition of a fine in an amount of TL 13 million ($9 million) against Turkcell, which has the largest number of users of any mobile provider in Turkey.

Additional sector specific legislation applies in the financial sector, under the Banking Law of 2005, which imposes a strict duty of confidentiality subject to criminal and civil sanctions, and the Bank Cards and Credit Cards Law of 2006, which restricts retention or re-use by retailers of customer information as a result of credit transactions.

Alas, at the same time, many sectors of the Turkish economy remain largely unregulated, including important government projects such as electronic health records, which contain highly sensitive information about patients' prescription histories and lack adequate measures of privacy and data protection. For example, the Turkish Ministry of Health established a comprehensive Medication Tracking System (''MTS'') for purposes such as ensuring drug safety, accounting for use of medications and engaging in pharmacovigilance. The MTS is based on a Ministry of Health regulation titled ''Medicine Tracking and Evaluation.'' Under this scheme, massive amounts of personal data, including name, age, diagnoses, hospital visited, medicines purchased, and adverse effects, are collected and stored in a centralized database. Not only health care practitioners but also pharmacists are able to query the database about a patient's prescription history, as well as additional information, such as address and social security payments. Even more sensitive health data are collected and stored by Turkey's Social Security Administration.

The Medicine Tracking and Evaluation regulation fails to clearly articulate: the purposes for collection of patients' data, roles and responsibilities of various actors involved, procedures for reporting adverse medicine effects to third parties, and scope of access authorizations. Moreover, the MTS appears to violate data protection principles such as proportionality and data minimization, collecting large amounts of personal data and making them available to numerous parties without apparent necessity. It is evident that only comprehensive data protection legislation can quench the thirst of public sector entities for additional data and subject such entities to requirements of data minimization, security, and retention limitation.

Moving Ahead

In order to give substance to the latest constitutional amendments and progress toward harmonization and partnership with the European Union, Turkey should adopt data protection legislation based on the fair information principles (''FIPs'') set forth in the 1980 OECD Privacy Guidelines. In addition, Turkey should establish a privacy and data protection authority charged with enforcing data protection legislation against both the private and public sector, in a manner commensurate with the structure of Turkey's constitutional and administrative law.

While procedures for implementation and enforcement diverge, there is broad global consensus concerning the substantive data protection principles, such as purpose limitation, proportionality, transparency, and security. Indeed, even the United States, which does not yet have across the board data protection legislation, has recently endorsed regulation based on the FIPs. In its Report on ''Commercial Data Privacy and Innovation in the Internet Economy: A Dynamic Policy Framework,'' the U.S. Department of Commerce states: ''To provide consistent, comprehensible data privacy protection in new and established commercial contexts, we recommend that the United States Government recognize a full set of Fair Information Practice Principles as a foundation for commercial data privacy.'' Turkey, which has already recognized the importance of protecting the right to privacy and individuals' data in its constitutional amendment and acceded to Convention 108, should likewise adopt legislation setting forth the FIPs as statutory requirements. The FIPs should apply to both private and public sector entities and be enforced by a dedicated data protection authority.

Turkey's data protection authority should be established in a manner commensurate with its constitutional and administrative law. Under a recent decree, Turkey subjected all independent regulatory bodies to

the supervision of government ministries.[7] While the EU Data Protection Directive mandates a supervisory authority acting ''with complete independence,'' independence should be balanced against enforcement powers and accountability. Turkey's legal system favors strong enforcement agencies exercising broad powers and that are accountable to the judiciary branch over weak independent authorities effectively acting as ombudsmen. The European Commission approved the ''adequacy'' of data protection frameworks in Argentina and Israel, despite these countries' data protection authorities being integrated into the executive branch.[8] A similar structure, where the data protection authority is subject to strictly financial and administrative oversight of the Ministry of Justice, should be considered in Turkey.

Conclusion

Despite constitutional amendments and the establishment of a strong regulator in the field of telecom privacy, Turkey continues to lack a comprehensive data protection framework in the mold of EU Member States. This imbalance, subjecting a specific sector of the economy to what has sometimes been considered heavy-handed regulation, while leaving other sectors without regulatory guidance or oversight, should be corrected. A data protection authority, established in the spirit of European regulatory agencies and empowered to enforce the law in all sectors of the economy as well as against the state, would correct this imbalance and propel the Turkish economy towards European standards.

---

[7] Independent bodies in Turkey attached to ministries after new decree law, Hürriyet Daily News, Aug. 22, 2011, http://www.hurriyetdailynews.com/n.php?n=independent-bodies-in-turkey-attached-to-ministries-after-new-decree-law-2011-08-22.

[8] *See, e.g.*, Commission Decision of 31 January 2011 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequate protection of personal data by the State of Israel with regard to automated processing of personal data, OJ L 27/39, 1 Feb. 2011, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:027:0039:0042:EN:PDF.

   ISSN 1538-3423

Location privacy challenges in geolocation standards: an interdisciplinary study

Maria Luisa Damiani, University of Milan, I
Colette Cuijpers, TILT, Tilburg University, NL

April 4, 2013

Abstract

Application programming interfaces compliant with the W3C geolocation API specification provides a simple set of operations for requesting geolocation services through the Web which greatly facilitates geospatial data collection. Importantly, the specification prescribes that conforming implementations must provide a mechanism that protects the user's privacy. In particular such mechanism should ensure that no location information is made available through this API without the user's express permission [11]. In this work we address the question on whether this privacy mechanism grounded on the informed consent privacy model is sufficient for collecting geospatial content in compliance with privacy laws. The question is of practical relevance as the use of geolocation standards in line with privacy regulations would ease the collection of geospatial content for example through the involvement of citizens (i.e. participatory sensing). In this paper we present an interdisciplinary analysis spanning across technology and law, and driven by an application case. Although we confine ourselves to consider European regulations, we believe that this study can be of more general concern. A preliminary version of this paper has been presented in [3]

1 Introduction

Geospatial data collection is an enabling factor in a variety of applications. For example, partecipatory sensing and volunteered geography applications target the collection of geospatial content through the direct involvement of citizens [8, 7]. Citizens can participate in different ways to geospatial data collection. For example, they can provide fine-grained knowledge of the territory and annotate shared maps, as in wikimapia[1]. Or can act as mobile sensors, for example to report the location of illegal dump sites to some organization which can verify the information and report it to the appropriate local authorities.

However, whenever geospatial content refers to individuals it cannot be collected without providing individuals with privacy guarantees in compliance with law. As an example, consider a participatory project aiming at studying the social habits in a city, for instance where citizens spend most of their time, which are the social services they use and so forth. Volunteers living in the city could be equipped with a smartphone and provided with a simple application which periodically transmits their location to an application server. Analytical methods can then be

[1] http://wikimapia.org

applied on the collected data to discover mobility patterns as in [13]. Even though this initiative is carried out for legitimate and valued purposes, i.e. improving the quality of social life, it remains the problem of how to practically deal with the collection of personal location data in compliance with data protection law. The problem stems from the fact that location represents personal data and as such can only be processed in compliance with data protection legislation which contains obligations for data controllers (e.g. application providers) and rights and guarantees for data subjects (i.e. users). For example, the processing of location data under EU law may only be allowed with prior consent of the data subject, where prior consent is only valid as a legal ground for processing personal data when consent is specific, freely given and based on information which is clear, comprehensive, understandable for a broad, non-technical audience and permanently and easily accessible.

In this work, we focus on a specific class of services relevant for the collection of geospatial content, in particular the geolocation services which can be requested through interfaces compliant with the W3C geolocation API specification. The W3C geolocation API (simply API hereinafter) is qualified as *W3C recommended standard* since May 2012 [11]. This API provides the abstract specification of a set of operations, which embedded in Web pages, enable to estimate the location of the website visitors pervasively, i.e. across indoor and outdoor spaces, and across urban and rural environments. The API specification is privacy-aware, in that it prescribes that conforming implementations must provide a mechanism to ensure that no location information is made available through this API without the user's express permission [11, 6]. Notably this API is supported by all major Web browsers. This means that individuals can be potentially tracked on a worldwide scale and pervasively, with little programming effort. Privacy safeguard is thus a critical issue.

Our research is driven by the following question: *can we rely on the privacy mechanisms offered by the API to collect individual location data, in compliance with data protection law?* The question has important practical implications as the use of geolocation standards compliant with privacy regulations would greatly facilitate the development of data collection applications. In this paper we argue that the privacy mechanism offered by the current API is not enough to comply with the European data protection law. Therefore additional measures need to be undertaken by the developers and this inevitably impacts on the complexity of the application. To support our argument, in what follows we present an application in which location data from a community of students are collected in a participatory way on Web. Based on this experience, we present an analysis of the legal aspects and discuss a possible extension of the API.

The remainder of the paper is organized as follows: Section 2 presents some background knowledge. Section 3 introduces the application case and the related privacy analysis. Section 4 outlines the proposal of an extended API. An alternative research direction is sketched in Section 5. Final considerations are reported in the conclusive section.

2 Background knowledge

Before presenting the key features of the W3C geolocation API, we overview geolocation techniques commonly available on popular devices, e.g. smartphones. Next, we briefly describe those aspects of the European data protection regulation that are relevant to understand the limitations of the API.

2.1 Location techniques and services

The location of a mobile device, i.e. a smartphone, can be estimated using a multiplicity of techniques [10]. Crucial for privacy is the distinction between handset-based and network-based solutions: in the former case the location is computed by the mobile device itself, in the latter by a third party, e.g. telecommunication operator. The most popular handset-based positioning technique is GPS. GPS provides worldwide coverage and supports a range of location services with accuracies that range from a few meters to a few millimeters. However, when integrated in mobile phones, GPS localization presents severe limitations i.e. it is power consuming, moreover the device must be located in a position in line of sight with at least 4 satellites for the location to be estimated. This means that GPS cannot be used inside buildings, underground and in the so called urban canyons [10]. In those environments network-based localization techniques are appropriate, such as cellular (GSM/CMDA) and Wi-Fi (802.11) based systems. Recent hybrid location systems integrate different technologies, e.g. cellular, Wi-Fi and IP-based positioning, to offer unprecedented opportunities in terms of location coverage and accuracy. This explains the increasing concern for business models offering location as a service. Location services providers (LP) include Google Location Services, Apple, Skyhook and Microsoft.

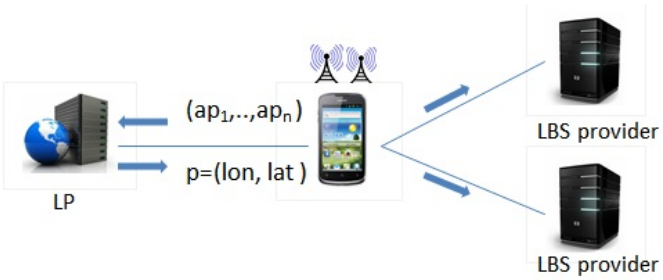

Figure 1: LBS Architecture: the client transmits to the LP contextual information (e.g. set of Wi-Fi access points, $AP_1, .., AP_n$) to obtain the location

Figure 1 illustrates the typical architecture of a location-aware application in which the location is computed by a LP. The client requests the location from the LP and then forwards such location along with the service request to, e.g. the LBS (location-based service) provider. Typically, LPs compute the location based on the contextual information sent by the client, e.g. Wi-Fi access points in proximity (e.g. $AP_1.AP_n$), by properly matching these patterns against a large and proprietary database of geo-referenced access points and cell towers. In an urban setting, with high density of Wi-Fi networks, the location returned by the LP can have an accuracy of 10-20 meters. Note that users can interact with diverse LBS providers offering different services, therefore LPs are in a location to compile extensive records of users' location and movement.

2.2 Location data protection in Europe

In view of the privacy issues raised in this paper, it is important to have some basic understanding of the general Data Protection Directive and the so-called ePrivacy Directive, the main pillars of the EU legal framework regarding processing of personal and location data.

2.2.1 Data Protection Directive (95/46/EC)

The general Data Protection Directive (hereafter: DPD) consists of a layered system of three levels. The first level is the general level that applies to all processing of personal data. The second level is applicable when sensitive data are being processed. The third level is applicable when personal data are being transferred to third countries. The layered system is cumulative, meaning that if sensitive data are being transferred to third countries, all three levels apply. The issues we address in this paper mainly concern the first level, therefore the second and third level will not be discussed. Besides the three levels within the DPD, the EU legal framework on data protection consists of two more levels of protection, including Directive 2002/58/EC (ePrivacy Directive).

The DPD applies to the processing of personal data which is defined as "any information relating to an identified or identifiable natural person (data subject)", while processing covers "any operation or set of operations which is performed upon personal data". Both concepts are interpreted in a very broad manner. As a main rule, the DPD stipulates that personal data "may only be processed fair and lawful". What fair and lawful entails, can be derived from the other provisions in the Directive. Main requirements: having a specific purpose and legitimate basis for the processing of personal data. Processing is only allowed in accordance with the specified purpose and may not go beyond this purpose. Regarding the quality of the data it is determined that data must be relevant, accurate, not excessive and up to date. Sound security measures also need to be taken in order to protect data from being corrupted or destroyed. Furthermore, the data controller has the obligation to inform data subjects (and in some cases the Data Protection Authority) regarding data processing. Data subjects have the right to access, rectification, erasure, blocking and the right to object. To ensure enforcement, Member States are obliged to put in place effective sanctioning mechanisms in case of infringement of the data protection rules.

The obligations in the DPD are addressed to the data controller, defined as the entity that "determines the purposes and means of the processing of personal data". In view of the actual processing, a controller can engage a processor, an entity "which processes personal data on behalf of the controller". However, the controller is responsible for and must put in place processing contracts with their data processors. As we will see in the legal considerations in Section 3.2.1, it can be rather difficult to identify the controller and processor(s) when the W3C geolocation API is used to provide a LBS. This is however very important, as it is the controller who must comply with all rights and obligations laid down in the legal framework.

2.2.2 The ePrivacy Directive (2002/58/EC)

The provisions of Directive 2002/58/EC (as amended by 2009/136/EC) particularize and complement the DPD in the field of electronic communications services. Importantly this Directive lays down rules regarding the processing of location data, in particular it states that: "location data may only be processed when made anonymous or with prior consent and only for the duration necessary for the provision of a value added service", being: "any service which requires the processing of traffic data or location data other than traffic data beyond what is necessary for the transmission of a communication or the billing thereof". The main difference between the two Directives concerns the fact that the regime to process location data in providing value added services is stricter than the general regime regarding the processing of personal data.

Prior consent and anonymisation are the only valid grounds for processing location data. In practice, the only valid ground in most cases will be prior consent as the Art. 29 WP has stipulated that "true anonymisation is increasingly hard to realise and (. . .) the combined location data might still lead to identification" [12]. Therefore, the core legal consideration addressed in Section 3.2.1 concerns the difficulties of the requirement that location data may only be processed with the prior consent of the data subject, i.e. user.

2.3 The W3C geolocation API specification

The W3C geolocation API provides the abstract specification of a few simple operations to request location services from inside a program. Notably, the API is agnostic of the positioning technology, namely the geolocation service can be requested without specifying which technology is to be used to estimate the location, i.e. either GPS or Wi-Fi based or IP based positioning. Operations can be embedded in Web pages, using a scripting language like Java Script. Moreover, the API is supported by all major Web browsers. This means that the visitors of a geo-enabled website can be localized across different operating systems. Note however that this simplicity of use is paid in term of flexibility, in that the application provider is not in the position of exercising any control on the way the location is computed, in particular on whether the location is computed by the client device itself or through a LP while this is important for privacy. What happens in practice is that the Web browser translates the API operations into location service requests for a LP[2]. Therefore, any time the user is located, such location is communicated to such LP. Note that the LP depends on the Web browser.

```
interface Geolocation {
  void getCurrentPosition(PositionCallback successCallback,
                          optional PositionErrorCallback errorCallback,
                          optional PositionOptions options);

  long watchPosition(PositionCallback successCallback,
                     optional PositionErrorCallback errorCallback,
                     optional PositionOptions options);

  void clearWatch(long watchId);
};

callback PositionCallback = void (Position position);

callback PositionErrorCallback = void (PositionError positionError);
```

Figure 2: Abstract specification of the geolocation operations

2.3.1 Technical features

For the sake of completeness, we briefly overview the key features of this API. The interface provides two main operations, called getCurrentPosition and watchPosition, to get the single location and to continuously update the location, respectively. Note that the latter operation allows users' tracking. The abstract specification of these operations is reported in Figure 2. Upon the location service request, the LP returns the following information. Some of these data are provided on developer's discretion (i.e. are optional):
- Timestamp of the location
- Position as latitude and longitude
- Altitude (optional)

[2]http://www.mozilla.org/en-US/firefox/geolocation/

- Accuracy of the location (Long, Lat) in meters
- Accuracy of the altitude (optional)
- Heading (the direction is expressed as angle with respect to North) (optional)
- Speed in meters per second (optional)

3 Application case

The application case consists of a Web application involving the participants of a seminar on location privacy[3]. The purpose of this application is threefold: i) to experience an active learning approach to location privacy based on user's participation; ii) to highlight location privacy issues related to the use of W3C geolocation API; iii) to collect data on the accuracy of the location service offered by popular LPs throughout Europe. This application has been developed in the framework of the EU project Modap (www.modap.org).

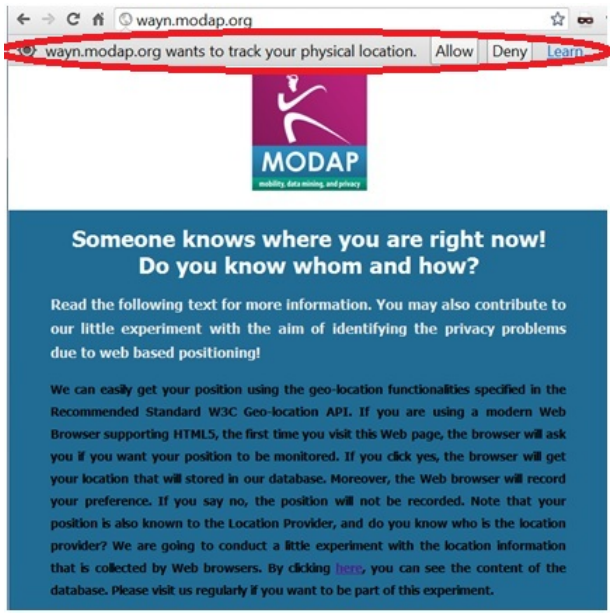

Figure 3: As the Web page is accessed the user is prompted with the request of consent (in red)

3.1 Location data collection through the Web

The Web application is conceptually simple: it computes the location of each user visiting the Web site from any device, mobile or not, and stores this information together with the user's IP and few additional information in a repository on the server. Users can then inspect the content of the repository.

The location is estimated as follows: first the application invokes the geolocation operation provided by the API (i.e. the getCurrentPosition operation). If the location is correctly returned, it is recorded in the repository along with additional information. Conversely, if the location is not available, because for example the user denies his/her consent to location collection,

[3]Summer School on Privacy-aware Social Mining, mss2012.modap.org

then the application computes a coarse location based on the IP address and using commercial datasets reporting the association between IPs and geographical locations. In this case, the location is determined locally. Note that the application does not include extra code for privacy support beyond what is provided by the API. The application, called WhereAreYouNow (Wayn) is available on http://wayn.modap.org. The participants of the seminar - more than 50 students (PhD and M.Sc students) from 13 countries mainly in Europe - were invited to visit the Web site in advance and leave their footprints.

3.1.1 Accessing the Web site

The interaction with the Web site is as follows: the visitor of the Web site, using for example Google Chrome, is presented with the page in Figure 3. The text explains the purpose of the data collection project and asks visitors to contribute to it. The first time the user visits the Web site, the Web browser asks the user to grant or deny his/her consent to the disclosure of location. Note that the interaction is completely transparent to the application. The user can then reply yes or no or even not reply. If the user replies yes then the Web browser records the preference, i.e. consent will not be requested again, unless the user deletes this preference, then forwards the geolocation request to the LP. Conversely, if the user denies consent or does not reply in a given amount of time (e.g. 10 seconds), our application records the IP and further information, e.g. time.

In addition, users are provided with a feedback on their location. Feedback is important to increase user's awareness and enhance user's experience. Feedback is provided by displaying on a map the logged locations along with the additional data stored in the repository. The map

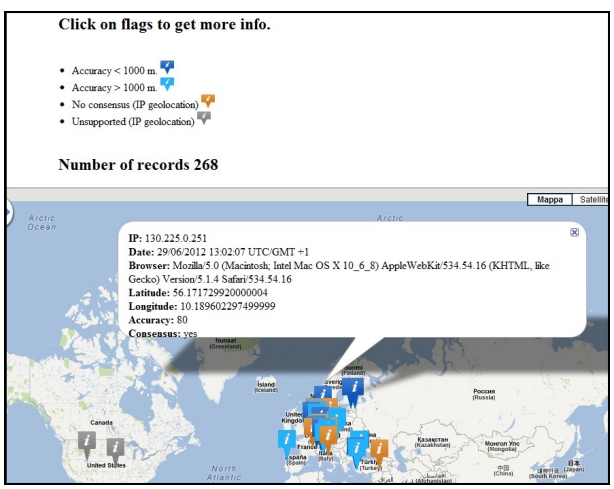

Figure 4: The map displays the content of the repository

in Figure 4 shows what the users can see, i.e. the content of the repository. This map can be accessed through the home page of the Web site. The map reports a pin for each record, i.e. visit. Pins are displayed in different colors based on the accuracy of the corresponding location. The most accurate are the locations identified by blue pins. Generally these locations are estimated based on either GPS or through Wi-Fi-based positioning. The sky colored pins are those typically detected by the LP based on the IP address. This happens for example when the user connects to the Web site using a cabled connection and no other geo-enabling device

is on (Wi-Fi or GPS). Conversely, if the user has denied consent or has not replied in the given amount of time the application displays the coarse location estimated based on the IP address and the pin is orange. Finally, gray pins are used in those situations in which the LP cannot compute the location, for example because of an error, or because the geolocation request is not accepted.

Each pin is shown along with the information recorded in the database, including IP address, time, coordinates, location accuracy and the Web browser footprint (i.e., Web browser configuration). One additional information that is currently not displayed to users, is the unique code, acting as pseudonym, stored as cookie on the mobile device. The pseudonym is used to correlate locations from the same user i.e. build the user's trace. If the user removes the cookie, a new pseudonym is created and thus also a new corresponding cookie. Note that this practice is left outside the legal analysis in this paper as currently a lot of debate exists within the EU regarding how to practically deal with the revised legal regime on cookies. However, the legality of this practice constitutes an interesting question.

3.1.2 Lessons learned

Before discussing the specific privacy issues which emerge from this experiment, it is worth mentioning interesting outcomes, of more general interest, of this application:

- *Creating privacy awareness in education*. This experiment allowed us to illustrate how location can be used to break the apparent anonymity of users and reveal personal details. For example, it may be relatively easy to discover where a person lives based on his/her trace and temporal information. For example, if an individual has accessed the Web site early in the morning from a residential area (the typology of area can be identified for example through StreetView), it is likely that the location is in proximity of user's home. Moreover, if the same person has left a footprint at the seminar location (we remind that locations of the same users can be correlated), the identification is fairly straightforward.

- *Access constraints to geolocation services*. A few participants tried to visit the Web site anonymously using the Tor onion routing system[4]. In that case, the geolocation request does not return a location, but an error. The gray pins displayed on the map in Figure 4 report points which are completely meaningless as the location is computed based on the IP of one of the nodes of the onion routing infrastructure. In essence the LPs are generally aware of the user's IP address.

- *Insights into the accuracy of geolocation services*. From this experiment, it turns out that the accuracy achievable from a LP is a few tens of meters uniformly across the different countries, such as Estonia, Turkey and Switzerland. Note however that official statistics from LPs are not available. Ideally, by promoting a larger scale participatory project, it would be possible to gain deeper understanding on the capabilities of third party geolocation services and come up with useful statistics. We leave this point for future work.

[4]https://www.torproject.org/)

3.2 Privacy issues

We turn to consider the privacy issues which emerge from this experiment. We recall that no additional support for privacy is provided beyond the basic mechanism offered by the API. Therefore users have only limited control over the disclosure of their location, as they can only accept or deny consent. Moreover, the experiment makes clear that users have insufficient information to express informed consent. We report some informal considerations here below. Then we analyze the question from a legal point of view.

- Users are typically not aware of the third party that behind the scenes computes the location, i.e. the LP. Even though certain Web browsers somehow provide this information, the actual implications are not clear to most of the users.
- As a consequence, the denial of consent is not correctly understood. Actually denying consent means that the location is not computed by the LP. This, however, does not prevent the application from computing the location in some other way, for example based on IP.
- Users are not aware of the location actually transmitted, unless using application-dependent functionalities, as we have seen before. Users are surprised of the high accuracy that can be achieved, whenever GPS is not used.
- Users are not aware on whether the location is just computed once or repeatedly, i.e. users are tracked. This because the current implementation of the standard does not make clear which geolocation operation is precisely used (i.e. getCurrentPosition or watchPosition, see Section 2.3.1). Therefore it may happen that the user believes to give his/her consent to the disclosure of the single location while in reality the user is tracked for the time the Web page is accessed.
- Users are not aware of the fact that by law they could request the removal of information recorded in the repository.

3.2.1 Legal considerations

In the introduction the question is raised whether we *can rely on the privacy mechanisms offered by the API to collect individual location data, in compliance with data protection law* In this section we assume the applicability of the ePrivacy Directive without entering into the discussion whether or not the ePrivacy Directive is applicable. For a deeper understanding of this very complex discussion we refer to [1]. Moreover, we discuss some of the legal considerations that are directly relevant in view of amending the API, while acknowledging a much wider array of legal implications in need of more in depth research.

We focus in particular on the difficulties regarding the legal requirement of prior consent when use is being made of the standard W3C geolocation API. Prior consent, as a legal ground for processing location data, is only valid when consent is specific, freely given and based on information which is clear, comprehensive, understandable for a broad, non-technical audience and permanently and easily accessible. In relation to geolocation services on smart mobile devices the Art. 29 WP has clarified the concept of consent: it cannot be obtained freely through mandatory acceptance of general terms and conditions, nor through opt-out possibilities; the

default should be that location services are 'OFF'; when switched on with consent, the user must continuously be warned that geolocation is switched on; data subjects need to renew their consent after changes have occurred in the service and even without changes reaffirm their consent at least once a year; besides the possibility to withdraw consent at any time, there must be a simple means, free of charge, to temporarily refuse the processing of location [12].

In view of the information to be provided in order to obtain valid prior consent the ePrivacy Directive requires information regarding the type of location data, the purposes and duration of the processing and whether the data will be transmitted to a third party for the purpose of providing the value added service. From the ePrivacy Directive it also follows that the processing must be restricted to persons acting under the authority of the provider of the public communications network or publicly available communications service or of the third party providing the value added service. Furthermore, processing must be restricted to what is necessary for the purposes of providing the value added service. From the foregoing three main legal considerations emerge regarding: *valid consent; restriction to persons acting under authority; and, restriction to what is necessary for the purpose*.

Valid consent. When use is being made of the W3C geolocation API Figure 3 shows that the user is informed that wayn.modap.org wants to track the physical location of the user. The user has the option to click the yes button, the no button, to click learn more or not to click anything at all. A user might expect that the *learn more* button will lead to information provided for by Wayn, but it leads to the privacy policy of the Web browser. Even though here some information can be found on things like withdrawal, the validity of the consent can still be questioned. For example is providing a *learn more* button that refers to a Web page of the Web browser "prior information provided for by the provider of the value added service"? And what about other characteristics such as the fact that consent is only requested once, and other legal requirements such as the need to make ongoing tracking visible for the user at all times. As the standard W3C geolocation API does not leave any room for Wayn to customize the location request, the only option for Wayn in order to comply with data processing legislation is to provide all information necessary in order to obtain valid prior consent on the Web site that appears with the location request. As the request appears in a separate dialogue box on top of the page, it is questionable if the user is completely aware of the relation between the Web content and the request (aggravated by the fact that the learn button leads to the Web browser). From the perspective of Wayn it might not be preferred to provide long lists of information on its Web page, or worse, Wayn might not have this information available as will become clear below. Moreover, the information will probably be rather confusing as Wayn will need to explain how other parties, such as the Web browser and the Location Provider are involved in the processing of personal and location data, while in fact Wayn has no control over this process whatsoever. This leads to the second legal consideration.

Restriction to persons acting under authority of. Even though we consider Wayn to be the controller as it is Wayn wanting to learn its users positions, from a legal perspective this qualification is questionable. It is not Wayn who determines the means of processing, but the Web browser/Location Provider. Wayn has no control over the type of technologies and data that are being used to pinpoint the users' location. Also, not Wayn but the user determines the Web browser. So, as Wayn cannot specify the type of positioning technology to use, including the specification whether the position should be determined locally or by a third party,

it might not even be possible for Wayn to provide users with all the information necessary to obtain valid consent. This means we encounter two legal problems. First, there is no possibility for Wayn to exercise any control over the way in which the user's consent is obtained. And second, Wayn might not even be capable to provide proper information regarding the data processing process, which is required in view of valid prior consent. So Wayn must rely on this information to be properly provided by the Web browser, which party is chosen by the user.

This leads to a whole array of interesting legal questions regarding the processing of personal and location data by the third parties involved within the process of providing LBS and the lack of clarity for the average user regarding the involvement of these parties. A user of a Web site expects to be dealing with the owner of the Web site when he is asked if his location can be pinpointed, while in fact it is his Web browser that acquires the location information and then passes this onto the Location Provider, a party probably not even known to the user. The LP then redirects the service request of the user, along with his location information, back to the Web site Wayn. So the Web browser as well as the LP are aware of IP address and (coarse, depending on whether or not consent has been given) location of the user.

If we look at the Firefox browser we can see that it presents a seemingly clear privacy policy (http://www.mozilla.org/en-US/legal/privacy/firefox.html): the Web browser collects the Wi-Fi info and forwards the request to the LP. So the Web browser is the intermediary between the user and Wayn, but involves the LP in this process. Moreover, as such the Web browser and the LP are aware of the information and thus in a position to process the personal and location data as well, even beyond the purpose of the provider of the LBS for whom the request is being "translated". Whether or not and how this is being done is unclear to us, but also to the user whose privacy is at stake. But even when these parties only process the data as far as is necessary in view of the location request of Wayn, legal problems remain. The qualification of the Web browser and the LP under the DPD and the ePrivacy Directive is problematic as they do not fall under the strict authority of Wayn. Moreover, no formal contracts exist establishing a legal controller - processor relationship between Wayn and the Web browser on the one hand and Wayn and the Location Provider on the other hand. What the implications are of this rather complex situation is in need of more in depth research, from a computer science as well as from a legal perspective.

Restriction to what is necessary for the purpose. In relation to the requirement that processing is allowed only if necessary for the purpose, a comment can be made relating to the characteristic of Wayn that it calculates and stores a coarse location position on the basis of an IP address even when consent is denied. This does not comply with data protection legislation. IP addresses are personal data and thus the processing of this data in combination with (coarse) location data is only allowed on the basis of a legal ground. Without consent the location may not be calculated.

4 Extending the API

To partly alleviate the lack of transparency that is one of the underlined limitations, we sketch a possible extension of the API. For the sake of clarity, we illustrate these additional features through the screenshots of a prototype. The prototype has been built by simply re-defining the API operations and encapsulating the native operations in a JavaScript object. This means

that the Web browser has not been extended with additional functionalities (i.e. plug-in) while the graphical interface handling the interaction with the user is not fully integrated. These limitations are not really relevant for the purpose of demonstrating the purpose of the extension. The extended API provides the following features:

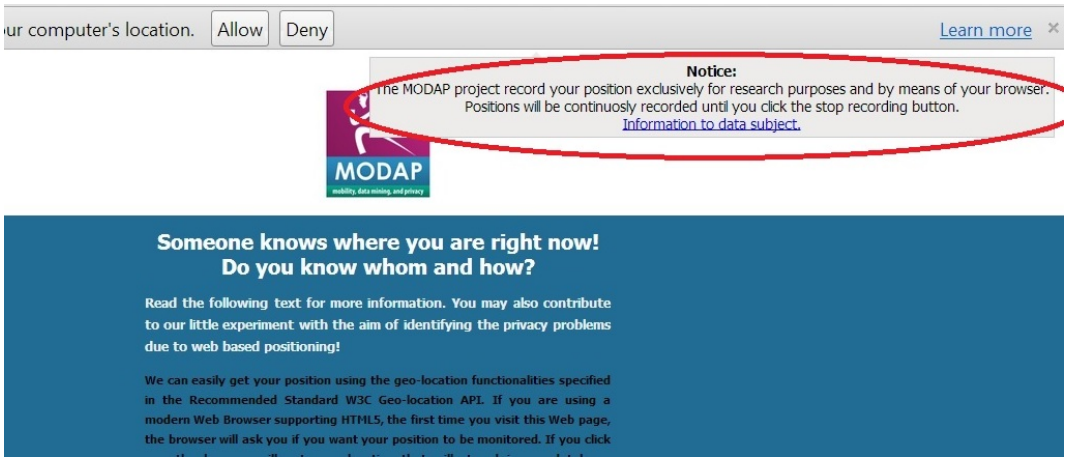

Figure 5: Displaying the application notice at the time the user's consent is requested

- The geolocation operations are extended with one additional parameter, the *url* of the privacy notice of the application. Even though the idea is simple, we believe that this is a substantial improvement in the direction of aligning data collection to data protection norms. The privacy notice reports key information, one of the obligations deriving from data protection regulation. An example of notice, reporting the purpose of the application together with a link to further details, is illustrated in the screenshot in Figure 5.

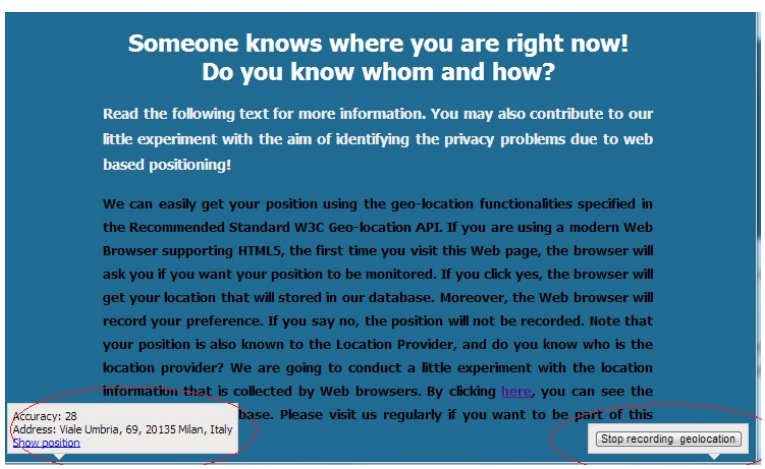

Figure 6: Accuracy of the location being recorded and indication of the location service being requested after the consent is granted

- Users are made aware of the fact that their location is either acquired once or conversely users are tracked. In the latter case the user can explicitly stop location tracking. As an

additional functionality, users get a feedback on the accuracy of the computed location. The interface is illustrated in the screenshot in Figure 6.

- Users can see the location information being collected by the application on a map. The map only display the trace of the specific user (not of the other users). In order to minimize the information being collected, non necessary information such as the IP address are no longer recorded. A sample map is reported in Figure 7

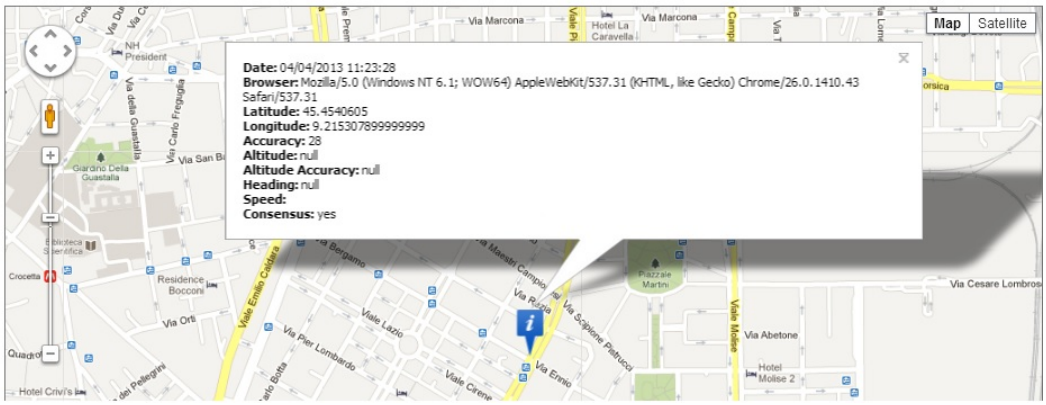

Figure 7: The map reporting the user's data being collected

5 An alternative viewpoint on the problem

The implicit assumption underlying the discussion on the API extension is that the LP is trusted and thus behaves in compliance with law. Accordingly the data protection goal is to ensure that users are aware of how their personal data are used. In this section we mention a different viewpoint on the problem according to which the LP is untrusted. Under this hypothesis the data protection goal is to minimize the transfer of data so as to reduce the need of trust. These two situations we have delineated, i.e. trusted vs. untrusted party, are at the basis of the concepts of soft privacy and hard privacy respectively [5]. While the solutions prospected in the previous sections can be categorized as soft privacy solutions, an emerging direction of research is towards a hard privacy solution.

5.1 Towards a hard privacy solution

A first approach to the problem is presented in [4]. The work is driven by the consideration that the amount of information that the user transmits to the LP exceeds what is really necessary to determine the users' location. In fact every time a service is requested from a given place, e.g. home, the client transmits the same or similar contextual information, e.g. Wi-Fi access points. Therefore if clients would acquire the capability of recognizing autonomously the places that have been already visited, the location information would only be requested to the LP when it is strictly necessary. As a result, the communication would be minimized. To implement this idea, we devise an approach based on the metaphor of *private place* [2]. Private place is an abstraction which conceptualizes the intuition that there are some regions of space that

belong to the personal sphere. The intuition is that whenever the user is in a private space, the location should not be disclosed to the LP. In order to recognize whether the position is inside or outside a private place, using a consumer device, without interacting every time with the LP we take inspiration from previous approaches such as [9] to develop a solution which associates every place a radio fingerprint, specified in terms of Wi-Fi access points. Private places are thus recognized by comparing the networking infrastructure detected in a point, e.g. the Wi-Fi access points, with the set of radio fingerprints.

Minimizing the interaction with the LP, however, does not forestall the disclosure of the private place to the application provider. Every time, the user requests a service from, say, home, where home is a private place, the position conventionally associated with the private place at the time the place is defined, is disclosed to the application provider (conversely the service could not be requested). Therefore if the application provider is untrustworthy or collude with the LP, location privacy is again at risk. To achieve a comprehensive protection of location from both the LP and the application provider, the approach is to use privacy rules. An example of privacy rule is the following:

$$Home, [19:00, 08:00] \rightarrow cityOf(Home)$$

$Home$ is the name of a private place. The rule means that when the user is at home during the night, the location communicated to the application provider is the city in which home is located. In [4] we show a possible integration of the system, called Placeprint, with the geosocial networking application Google Latitude.

6 Conclusions

This work focuses on the compliance of the W3C geolocation API with European Data Protection norms. To deal with the problem we adopt an interdisciplinary approach which seeks to mediate between the mindset and the competences of computer scientists and legal experts. Importantly the legal analysis has brought to the forefront an array of questions that are worth being addressed as part of future work. Moreover this work also suggests novel directions of research towards privacy-enhancing technologies ensuring stronger protection from untrustworthy third party location providers.

Acknowledgements The work reported in this paper has been partially supported by the EU FET-OPEN project MODAP. We are grateful to Yucel Saygin, who supported the development of the Wayn case study and Vincenzo Pupillo who developed the software.

References

[1] A. C. Cuijpers and M. Pekarek. The Regulation of Location Based Services: Challenges to the European Union Data Protection Regime. *Journal of Location Based Services*, (5)3-4:223–24, 2011.

[2] M.L. Damiani. Third party geolocation services in LBS: privacy requirements and research issues. *Transactions on Data Privacy*, 4(2):55–72, 2011.

[3] M.L. Damiani and C. Cuijpers. Privacy-aware geolocation interfaces for volunteered geography: a case study. In *Proc. ACM SIGSPATIAL Workshop GeoCrowd*, 2012.

[4] M.L. Damiani and M. Galbiati. Handling user-defined private contexts for location privacy in LBS. In *Proc. ACM GIS '12*, 2012.

[5] G. Danezis. An Introduction to Privacy Technologies. http://www.cl.cam.ac.uk/teaching/1112/SecurityII/2012-security2-4-danezis.pdf,2012.

[6] N. Doty, D. Mulligan, and E. Wilde. Issues of the W3C Geolocation API. Technical report, UC Berkeley, School of Information, 2010.

[7] D. Estrin and K. M. Chandy et al. Internet predictions. *IEEE Internet Computing*, 14(1):12–42, 2010.

[8] M. Goodchild. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69:211–221, 2007.

[9] J. Hightower, S. Consolvo, A. LaMarca, I. Smith, and J. Hughes. Learning and recognizing the places we go. In *Proc. of the 7th international conference on Ubiquitous Computing*, UbiComp'05, 2005.

[10] E. De Lara, A. Lamarca, and M. Satyanarayanan. *Location Systems*. Morgan & Claypool Publishers, 2008.

[11] W3C. Geolocation api specification. http://dev.w3.org/geo/api/spec-source.html, 2012.

[12] Article 29 Data Protection Working Party WP185. Opinion 13/2011 on geolocation services on smart mobile devices. http://ec.europa.eu/justice/policies/privacy-/docs/wpdocs/2011/wp185_en.pdf.

[13] Y. Zheng, L. Zhang, Xie X, and W-Y Ma. Mining interesting locations and travel sequences from gps trajectories. In *WWW*, 2009.

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*

[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.

meaning of Article 27 may be a useful instrument for providing guidance as to the ways in which 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*

---

[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.

individual".[12] To conclude the element of result is met when the use of the data is: *"likely to have 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.

---

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

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 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

---

[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.

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 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.

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.

---

[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 (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

---

[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)

relationship 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] 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]

---

[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

[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).

privacy enhancing 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.