A Privacy Preserving Design Framework in Relation to an Environmental Scanning System for Fighting Organized Crime

Anne Gerdes

A Privacy preserving Design Framework in relation to an Environmental Scanning System for Fighting Organized Crime

Anne Gerdes

Department of Design and Communication
University of Southern Denmark, Kolding, Denmark
gerdes@sdu.dk

Abstract. This paper represents preliminary recommendations regarding the development of a privacy preserving system design framework related to the EU project, ePOOLICE, which aims at developing an environmental scanning system for fighting organized crime by providing law enforcement agencies opportunities for strategic proactive planning in response to emerging organized crime threats. The environmental scanning is carried out on a variety of sources, focusing on early warning and the disclosure of crime trends, not on individuals. Consequently, personal data are not relevant in the information context of ePOOLICE, and therefore the system will not make use of any kind of sensitive information. Particular attention are paid to the environmental scanning of data streams from social networking sites; based on the assumption that ethical and privacy issues with regard to social media scanning represent a significant challenging scenario to meet in developing a privacy preserving framework for ePOOLICE.

Keywords: Privacy preserving design, engineering ethics, environmental scanning, open source intelligence, organized crime fighting.

1 Introduction

The ePOOLICE project[1] aims at developing an efficient and effective environmental scanning system as part of an early warning system for the detection of emerging organized crime threats and changes in modus operandi, focusing on e.g., illegal immigration, cocaine trafficking, human trafficking and cybercrime. In the ePOOLICE system the environmental scanning is carried out on data streams from a variety of sources, counting both open sources (e.g.; Web resources, news, statistics, libraries,

[1] Project full title: "early Pursuit against Organized crime using envirOnmental scanning, the Law and IntelligenCE systems". Grant agreement no: 312651, THEME [SEC-2012.6.3-1]:[Developing an efficient and effective environmental scanning system as part of the early warning system for the detection of emerging organised crime threats - Capability Project].

research reports, and social media) and restricted sources (e.g.; hospital statistics, border traffic statistics, financial transaction statistics, narratives from police districts). Consequently, different kinds of public online data streams, including social media data streams, feeds into the system's knowledge repository that provides a rich taxonomy of domain knowledge. Moreover, data processing is facilitated by means of data analysis techniques, enabling the extraction of descriptive and predicative meanings used for inferring hidden states, i.e.; weak signals or indicators of organized crime activities. For example, one type of indicator may be an indicator of another type of organized crime; hence organized cannabis growing also indicates human trafficking (manpower for cannabis farming).

ePOOLICE operates at the strategic level of open source intelligence and takes into account modus operandi and crime trends. Consequently, personal data are not relevant in the information context of ePOOLICE, and therefore the system will not intend to make use of or aggregate personal data or maintain a database for storing or managing personal data or other kinds of sensitive information.

The environmental scanning in ePOOLICE is conducted via a broad-spectrum scan of open sources; the system functions as a tool for tactical planning focusing on modus operandi, hotspot locations, crime patterns and (mega-) trends. Hence, the use context of ePOOLICE is situated at the strategic level, implying that the system does not support the operational level at all, but serves a pure preventive purpose in scaffolding sense-making activities carried out by law enforcement agents and analysts engaged in countering threats and acting proactively in dealing with upcoming trends in organized crime.

In this paper, particular focus is on threats to privacy occurring through the use of open source material from social networks; followed by an outline of preconditions for a privacy preserving design framework to be implemented in dealing with privacy issues raised by system functionalities. The reason for paying special attention to social media scanning is motivated by the fact that social networking platforms contain a vast amount of personal and or sensitive information and are typically conceived of as trusted personal networks among friends. As such, social media scanning represents a particular challenging topic to deal with in developing a privacy preserving framework for ePOOLICE. Hence, it is assumed that the type of potential ethical and privacy issues that may occur in relation to social media scanning are likely to occur in a weaker version in the scanning of other sources (e.g.; news, statistics, university research reports, demographic data). It is important to stress that all though the focus of this report is on social media scanning, this position does not entail the negligence of potential different types of privacy issues related to the scanning of other types of sources, which may require different means or techniques to solve.[2]

1.1 Purpose and Scope

With the above mentioned in mind, the purpose of this paper is to clarify and outline:

[2] For example, privacy preserving methods suitable for social networking data streams would not be suitable for trajectory data.

- ethical and privacy issues in relation to environmental scanning of open source data streams in the shape of social networking sites.
- an overview of privacy preserving data mining techniques, which may be applied in ePOOLICE in order to deal proactively with privacy protection in the context of environmental scanning of open source documents from social networking platforms.
- a preliminary privacy preserving system development design framework, which meets the threats of privacy infringements or violations without hindering the knowledge discovery opportunities of the system and the end user.

1.2 Paper Structure

Section 2 represents an overview of ethical and privacy issues related to the environmental scanning of open source data streams, specifically emphasizing data streams from social networking sites. Section 3 presents an overview of the Privacy by Design paradigm and privacy enhancing methods of relevance for the development of a preliminary privacy preserving design framework. As the project progresses, this framework will be refined continuously in order to be able to balance data privacy and data utility by carrying out privacy impact assessments of tools, processes and methodologies developed. Section 4 summarizes the current findings.

2 Ethical and Privacy Issues in Relation to Environmental Scanning of Data Streams from Social Networking Sites

Advancements in data mining techniques used on social networking platforms have exacerbated privacy concerns. Gradually, individual citizens have become more and more transparent to a variety of actors and at the same time they have experienced a reduction in transparency with right to knowledge of what is being known about them, where and by whom. On top of this, as Web users, we contribute to our own potential de-privatization by spreading information about ourselves on the Web, i.e., by being present at social networking platforms, or by enjoying the convenience of seamless internet transactions based on personalized services in exchange for personal data. Needless to say, that this might raise privacy concerns associated with the lack of autonomy in controlling the flow of information about oneself across different contexts, as well as lack of confidentiality and trust in relying on that intended or unintended information-based harm will not occur.

From a legal point of view, personal information in social networking platforms is protected by the EU Data Protection Directive (Directive 95/46/EC). The requirements of this directive with right to the anonymity of the data subject is meet in the design intentions behind ePOOLICE, since the environmental scanning of social networking sites provides a systematic approach for exploring and mapping patterns of communication and relationships among networks at a general level without singling out actors, i.e. unique data subjects. Consequently, within the overall framework of the ePOOLICE project no data subject is identified or under surveillance, and no personal and intimate information *per se* is involved in the identification of relevant data and interpretation of relevant patterns of information and knowledge. But since personal (and often also sensitive) information is highly accessible online, the inherent risk of unintentionally identifying data-subjects during the raw data scanning process of social media is fairly high. Here, we have to bear in mind, that personal data includes information, which may identify an individual indirectly by means of different fragments of sources. Furthermore, the environmental scanning may come up with patterns of information and point to indicators that hold the potential to sort groups by race, belief, gender or sexual orientation, etc. Still, when based on objective statistical analysis, the use of criminal profiling by law enforcement agencies is legal. Nevertheless, following the precautionary principle, we need to stress the importance of avoiding potential discrimination, which affords categorization of people into damaging stereotypes.

Even though environmental scanning may slip under the radar of legal privacy restrictions, social media scanning may still imply privacy discomfort among people, due to privacy concerns regarding information traffic across contexts representing distinctive spheres in life. Users on social networking platforms are typically aware that data shared on social media (such as Facebook and Twitter) resides in a public or semi-public sphere. Applying privacy settings may well decrease the group of people with access to your data, but still not hinder that data are spread to others. But we cannot per default presume that people, when engaging in producing and sharing online content on social media platforms (Facebook, Twitter or blogs) do not have any expectations of privacy. All though users are aware that data are to some extent public it is reasonable to assume that they probably do not expect their online content to be made available as raw data set for environmental scanning. Consequently we may refer to *informational privacy* as individuals' ability to control the flow of personal information, including how information is exchanged and transferred [14].

A justificatory conceptual framework, for the systematic exploration of people's reactions to technology can be found in Nissenbaum [9], who has coined the term "contextual integrity" in order to explain for and tie adequate protection against informational moral wrongdoing. According to Nissenbaum, information flows always have to be seen in relation to context-sensitive norms, representing a function of: (1) the types of information in case, (2) the respectively roles of communicators, and (3) principles for information distribution between the parties. Consequently, contextual integrity is defined, not as a right to control over information, but as a right to appropriate flows of personal information in contexts with right to two norms [9: 127 ff]: Norms of "appropriateness" and norms of "distribution", i.e., the moment of transfer of information from part X to$Y_{1\ldots n.}$ Violations of one of these norms represent a privacy infringement [9].

In the case of ePOOLICE, new flows of information are established and may cause a potential violation of contextual integrity, since information gathering via environmental scanning of communication streams on social networking sites may possibly be judged inappropriate to that context and violate the ordinary governing norms of distribution within it. Likewise, the width spread scope of the scanning may raise concern among citizens. Here, we have to strike a balance, which can ensure that the infringement of privacy is minimally invasive for human rights without obstructing the opportunities for the intended aim of open source intelligence.

In ePOOLICE, social media data streams feed into the system's knowledge repository that provides a rich taxonomy of domain knowledge and facilitates data processing through fusion techniques[3], which enable the extraction of descriptive and predicative meanings used for inferring hidden states, i.e.; weak signals or indicators of organized crime activities. All though a human analyst evaluates all system output, a risk still exist that the knowledge discovery techniques to disclose emergent trends in social network data streams may deduce patterns that could be sensitive. Also, the aggregation power and scope of ePOOLICE facilitates analysts' reasoning, but at the same time requires that the analyst is particular careful in the evaluation of possible emergent patterns and predictive meanings in order not to get "carried away" by the knowledge discovery power of the system.

To summarize, in the development phase of ePOOLICE, the privacy preserving framework of ePOOLICE must balance data privacy and data utility in addressing:

1. the risk of unintended identification of individuals (data subjects) during the social media scanning.
2. the risk of disclosure of sensitive patterns due to the data analysis techniques applied.
3. citizens' surveillance concerns raised by social media scanning.

Points 1 and 2 are situated at the level of technical system development, whereas point 3 resides at the societal level. Also, but beyond the scope of this paper, the institutional and organizational level have to be dealt with in order to address ethical and privacy issues, which may arise in the use context of ePOOLICE. The abovementioned list will be extended during the project life time with the aim of systematically identifying and meeting future challenges as the development of ePOOLICE progresses (see www.ePOOLICE.eu).

3 Overview of the Privacy by Design Paradigm and Privacy Preserving Methods

The concept *Privacy by design* (PbD) was coined by the Canadian information and privacy commissioner Ann Cavoukian [2]. The philosophy behind PbD stresses the importance of pro-actively building privacy into the design, operation and management of information processing technologies and systems. To accomplish this, the following 7 fundamental principles are listed [2]:

- "Proactive not Reactive; Preventative not Remedial

[3] Fusion techniques use converted data from diverse sources, such as multi-source and multilingual information, which includes credibility assessment, with the purpose of disclosing patterns or correlations.

- Privacy as the Default
- Privacy Embedded into Design
- Full Functionality – Positive-Sum, not Zero-Sum
- End-to-End Security – Lifecycle Protection
- Visibility / Transparency
  Respect for Users."

These principles, taken together with the legal requirements, will form the basis of the privacy preserving framework of ePOOLICE.

As such, ePOOLICE has to be developed in legal compliance with EU member states privacy legislation. At the international level, the European Convention for the Protection of Human Rights and Fundamental Freedom (ECHR), which the European charter of Human Rights is based on, feeds into the local laws of EU member states and stresses the importance of the citizens' right to privacy and protection of personal data. Likewise, the Code of Fair Information Practices (which dates back to the 1973 report from the *US Secretary's Advisory Committee on Automated Personal Data Systems*, the U.S. Department of Health, Education and Welfare) expresses five fundamental principles of data management and record keeping, which still encapsulate the core of subsequent privacy guidelines. Hence, the Convention for the Protection of Individuals with right to Automatic processing of Personal Data (Council of Europe, 1981) positions data protection as a fundamental right, subsequently backed up by the Data Protection Directive (Directive 95/46/EC), to which member states national legislations are aligned. This directive is currently undergoing transformation and the status of the new directive is not yet settled. Also, the non-binding OECD Guidelines of Protection of Privacy and Transborder Flows of Personal Data (1980, revised in 1999) codifies eight internationally agreed upon principles related to fair information practices. Main elements in these FIPs guidelines and regulations are:

- The principle of fair and lawfulness – the processing must be legitimate and pursue specified purposes; data must be adequate, relevant, not excessive; accurate and up to date, and not stored for longer than necessary
- The principle of processing legitimacy: stresses the necessity to obtain the data subject's consent – some exceptions do exist (exhaustively listed)
- The principle of prohibition to process sensitive data (unless exceptions exhaustively listed: explicit consent of the data subject, if the data have been made public by the data subject)
- The principle of the rights of the data subject: including rights of access, rectification, erasure; to opt out, to not be subject to a decision based on some automatic processing)
- The principle of confidentiality and security of the data processed
- The principle of obligation to notify the supervisory authority (declaration or authorization)
- The principle of accountability of the data controller
- The principle of adequate level of protection in case of data transfer to third countries

In combination with legislation of member states, Principles of fair information practices (FIPs) and the PbD principles provide a useful guide for the development of a privacy preserving framework, which may be achieved by building the FIPs and the PbD-principles into all relevant phases of the ePOOLICE system's life cycle.

3.1 Preliminary outline of a privacy preserving framework and privacy preserving methods

A privacy preserving framework must be set up for in order to inscribe privacy in all design and developmental phases of of ePOOLICE. In what follows, I intend to elaborate on how to meet privacy challenges relevant in the context of social media scanning by addressing the above mentioned 3 points summarising privacy risks (section 2) and by briefly pointing to the PIA methodology.

3.1.1 Addressing the risk of unintended identification of data subjects during the raw social media scanning

The environmental scanning system moves across and filters all potentially relevant information sources and maintains and "learns" which patterns to look for, where to look and how reliable the source is, and mark the credibility of the information. In order to focus the scanning process, the environmental knowledge repository (EKR) represents a dynamic ontology of domain knowledge. Hence, in the EKR, domain knowledge is going to be stored temporary with the risk that data fragments might facilitate identification of data subjects. Therefore, use of filter techniques for hiding data or knowledge patterns, which disclose data subjects must be in place. The scanning is further facilitated by fusion techniques (a fusion tool box) used for knowledge discovery with the purpose of forming hypotheses about patterns of weak signals suggesting emergent trends in organized crime.

ePOOLICE does not collect intelligence to be used at an operational level for finding or monitoring individual suspects. Still, issues of potential misuse have to be dealt with, especially in the case of social media scanning, since it is unavoidable that the raw data streams from social media scanning will include personal data about both direct (i.e. name, address) and indirectly identifiable individuals (i.e. photos). Here an attack model representing a scenario of the aim of a malicious individual or party may inform system design with the aim of protecting data from any kind of misuse - such as "mission creep", implying that data are used for other purposes than the purpose for which they were collected, or privacy breaches violating individuals' privacy. Therefore, the system developers must define categories of disproportionate analytical queries, from which alert-systems may be designed such as to automatically block processing in case of attempt of illegal processing. Also, system functionalities must be in place in order to permit management of different user-levels of authorization. Consequently, logging of system access and processes applied may ensure tracking of possible unauthorized use of data and further support control procedures by internal data controllers as well as by independent authorities.

Anonymization techniques have to be brought to play to ensure that no personal or sensitive data are available in the data material from social media scanning [3]. According to the EU data directive 1995/46/EC, all means likely reasonable to be used either by the data controller or by any other person to identify a said person should be taken into account in order to determine if a person is identifiable. Hence, to minimize the risk of identifying a person, anonymization techniques may be applied to obtain individual privacy, but this is not a trivial task. As such, simply removing direct identifiers is not sufficient to yield anonymity. For instance *k-anonymity* hides a unique individual among k-1 others, thereby disclosing the real data by transforming them in a way which makes impossible the re-identification under a fixed anonymity threshold. But this technique does not provide privacy in cases where sensitive values in an equivalence-class lack diversity [8] – for instance data regarding age, health and zip-code may for a given class be equivalent and thereby unintentionally support re-identification. Hence, examples of re-identification from seemingly anonymous data have been described. As an example, Sweeney managed to re-identify the medical record of the governor of Massachusetts [13]. Also, the AOL search record of a user linked to a photo re-identified the person [1].

Obviously, anonymization in the context of ePOOLICE is challenging. However, to safeguard the system against misuse, which may reveal the identity of individuals, ePOOLICE must take steps to apply a privacy preserving framework, which allows for the anonymization of data and the build in of alert and blocking systems, which restrain potential misuse of data.

3.1.2 Addressing the risk of disclosure of sensitive patterns due to the data analysis techniques applied

Fusion techniques in ePOOLICE are used for inferring hidden states with the purpose of predicting emerging trends in organized crime (OC). Hence, in applying the knowledge discovery technique referred to as formal concept analyses, the aim is to discover as many signals of indicators as possible for new OC treats. Also, the idea of primary and secondary indicators will be qualified by the notion of temporal proximity, i.e. by paying attention to primary indicators as those associated with emerging or ongoing OC, considered short range radar detection, which equals detection monitoring. On the other hand, secondary indicators are associated with the opportunity for OC before it occurs, i.e. facilitators for new OC threats and part of a long radar detection system, which equals situation monitoring.

In order to enhance the recognition of weak signals and predictive meanings, the fusion tool box in ePOOLICE employs “soft fusion” techniques, which combine automated fusion and human fusion processes, thereby enabling utilization of experts abilities to interpret signals and identify complex patterns. Furthermore, decentralized fusion is introduced to support modular solutions, whereby complex fusion systems can be build and distributed over multiple devices. Environmental scanning, facilitated by such fusion products, provides the system with the kind of knowledge discovery power, which enables it to spot emergent patterns and weak signals of OC. As such, the system extracts predictive meanings, which exceed the scope of the original data set. These findings are subsequently presented to and evaluated by the analyst in order to scaffold further analysis and the forming of hypotheses.

Issues related to the use of fusion to infer hidden states raise serious concerns regarding surveillance since the linking of different forms of data from the scanning of social network interactions may possible be invasive to the privacy of individuals or groups. Moreover, social networking data are semantically rich, which makes anonymization a demanding task in the first place, since the extra semantics facilitates linking such data to background knowledge with the risk of disclosing sensitive data. Hence, knowledge discovery tools, i.e.; fusion techniques for linking data, infringe upon individuals and groups control of how information relating to them, their interactions and their networks are represented, and subsequently what patterns of predictive meanings can be inferred from such fused information.

With the development of advanced fusion techniques in ePOOLCE, the inference problem increases, i.e., the problem known from challenges in securing databases from having users to pose queries and deduce unauthorized information from the legitimate responses they get from the database [15]. Taking into account the concept of contextual integrity, we may find parallels between the inference problem and the possible treats to privacy from social media scanning and applied fusion techniques in ePOOLICE. For example, analysts may infer sensitive hypotheses, implying the potential identification of an individual. Consequently, to ensure that sensitive hypotheses cannot be deduced from social media scanning, methods of inference control must be taken into consideration.

To sum up, the development of the system must take proper steps to ensure that every phase in the data processing – all way through from the handling of raw social media data to the final extraction of predictive patterns - avoid identification of data subjects. Likewise, the challenge that reverse engineering techniques could disclose data subjects must be addressed in a way that balance privacy and traceability, since, in some situations, tracking is necessary in order to evaluate the credibility of system output. Hence, to a restricted group of users (data controllers) this option must be available even at the risk of disclosing fragmented data, which may identify data subjects.

3.1.3 Addressing citizens' surveillance concerns raised by social media scanning

Privacy issues in ePOOLICE may be adequately dealt with from a legal perspective and still yield privacy concerns due to the fact that alterations in flows of information may lead to violation of contextual integrity (see above, section 2). As mentioned above, users on social network sites may feel intimidated from learning that their online interactions serve as raw material for social media scanning. Hence, the overall judgment of ethical implications related to ePOOLICE goes beyond the scope of a standalone privacy evaluation of the system, implying that the context-sensitive tradeoff between privacy and security and citizens' right has to be taken into consideration as well.

From a public point of view, an example of European citizens' opinion on privacy and security issues can be found in a participatory technology assessment, which concludes that citizens are open to legitimate security measures for crime prevention, whereas reference to terror treats does not justify privacy limitations for most citizens [12: 26 ff]. Consequently, it seems to be the case that people are prepared to value security over legitimate restrictions of informational privacy in specific contexts reflecting individual dimensions of security. To elaborate on this from a legal point of view, any limitation to fundamental rights of privacy and personal data protection has to respect some basic principles in order to be legitimate and ensure that privacy is not violated. As such, limitations have to rest on a legal basis and must be formulated with such a degree of precision that it enables citizens to understand how their navigation and conduct in society are affected by the given limitation. Moreover, a restriction must pursue a legitimate aim, i.e., be in accordance with listed legitimate aims, formulated within each article of rights in the ECHR, as aims that justify interference. Furthermore, any limitation must correspond to a real need of society and must be seen as an efficient instrument (for instance in relation to crime reduction and security). Finally, the principle of proportionality seeks to guarantee that the limitation is balanced to the aim pursued. In order to minimize the infringement of privacy rights and to assess the proportionality of a restriction, the main issues to settle are whether the overall effect of the constraint is reasonable and whether it is the least intrusive mean available. Here, to ensure that privacy is not violated, the ePOOLICE project must see to that the requirement of proportionality of the privacy restriction is satisfied. Given these circumstances, the ePOOLICE project strives to enhance both privacy and security by introducing pro-active privacy preserving design principles throughout all stages of the development process. In this way, the project seeks to develop technological solutions that support privacy compliant use.

Yet, a problem still resides in the fact that an assessment of proportionality is not easy to deal with in a precise manner. Judging whether the privacy interference caused by ePOOLICE is a suitable, necessary and adequate mean for fighting organized crime on a strategic level, implies, among other things, a measurement of security gains. But, security advantages are not easy to calculate – neither ahead nor ex-post. Hence, from the fact that security technologies have proved to be effective, we cannot presuppose this outcome for ePOOLICE in advance. Also, if it turns out to be the case, ex-post, that we observe a decline in organized crime after the implementation of ePOOLICE, we still need to carry out a thorough evaluation to justify if and how ePOOLICE contributed to this outcome.

Nevertheless, in the case of ePOOLICE, and taking into consideration the recent focus on intelligence surveillance (the NSA surveillance case), we need to emphasis the importance of transparency. In general, when dealing with intelligence, citizens express concerns about the fact that they do not have insight in how information about them is used and for what purpose. During the system development phase, it is of outmost importance to find ways to enter into dialogue with European citizens in order to disseminate awareness about the ePOOLICE project.

3.2 General Recommendations for Setting up a Privacy Preserving Framework

In order to fully integrate privacy by design per default into the design process, the ePOOLICE project has established an ETHIC-TECH team consisting of consortium members (system developers and members with legal and ethical expertise)[4]. Accordingly, it is of outmost importance to make room for dialogue between engineers and experts in law and ethics in order to enact ethics during the system development process. Understanding the ethical impact of systems in context implies engaging in what Nissenbaum - who together with Friedman [5], is one of the leading figures in the field of value sensitive design (see for instance: [4], [10]) - coins “engineering activism” [11], and Van Hoven addresses as “front-loading ethics” [16] in proposing a pro-active approach to bring ethics to design. This ideal of activating ethical expertise in order to improve system development is encapsulated in the below mentioned quotation by Nissenbaum [11]:

“Humanists and social scientists can no longer bracket technical details - leaving them to someone else - as they focus on the social effects of technology. Fastidious attention to the before-and-after picture, however richly painted, is not enough. Sometimes a fine-grained understanding of systems - even down to gritty details of architecture, algorithm, code, and possibly the underlying physical characteristics - plays an essential part in describing and explaining the social, ethical, and political dimensions of new information technologies.”

Consequently, the overall objective of the ETHIC-TECH team is to scaffold activities within the privacy preserving design framework of ePOOLICE, with the purpose of striking a balance between data utility and data privacy. Hence, the privacy preserving framework is implemented in the scientific and technical coordination of the project to ensure that the project work flow of ePOOLICE incorporates a robust practice, which integrates ethical evaluation of tools, processes and techniques developed and or applied during the system development process. This is done with the aim of identifying and countering privacy threats without diminishing the knowledge discovery opportunities, which are essential to the project. Moreover, regular privacy assessments (PIAs) of potential privacy risks must be carried out in order to minimize negative privacy impacts in deliberation with relevant stakeholders [6]. In ePOOLICE we will make use of the data protection impact assessment (DPIA), which has been defined by the European Commission as a tool for evaluating potential privacy risks in relation to data processing[5]. For the time being, there is no standard to carry out PIA or DPIA since different methodologies coexists. Hence, in order to build in specific methodologies for ePOOLICE, in addition to risk methodologies, we will take

[4] Also, an external ethical advisory board with European experts on legal and ethical issues is in place.

[5] Commision recommendation, March, 9th, 2012, regarding preparations for the roll-out of smart meterings systems (2012/148/EU), §I, 3 (c.), See: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:073:0009:0022:EN:PDF

inspiration from PIA-methodologies, as well as the methods used in the security technology project VIRTOUSO [7].

4 Concluding Remarks

In this paper, special attention has been paid to ethical and privacy issues related to social media scanning due to the fact that this type of environmental scanning is taken to be particular demanding in connection to the ePOOLICE project. All though personal data are not relevant and no database are maintained for storing or mining personal and or sensitive data, privacy issues still arise. Hence, the risk of misuse, the risk of unintended identification of individuals and the risk of disclosure of sensitive patterns of information must be dealt with adequately by inscribing privacy preserving techniques into the design of ePOOLICE. Furthermore, due to citizens' general surveillance concerns related to intelligence, it is imperative to promote transparency about the ePOOLICE project.

The implementation of privacy protection in the design is not a trivial task. Therefore, a privacy preserving framework shall incorporate ethical evaluation of tools, processes and techniques developed all through the system development process. In accordance with the Privacy by Design philosophy, it is important to balance data utility and data privacy in order to meet privacy threats without obstructing the knowledge discovery opportunities in ePOOLICE.

Acknowledgments

The author would like to thank Estelle De Marco, Inthemis Managing Director, for fruitful collaboration on the ePOOLICE project as well as for valuable comments, which helped shape this article.

Research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 312651.

References

1. Barboro, M., Zeller T.: A Face Is Exposed for AOL Searcher NO. 4417749, The New YorkTimes, http://www.nytimes.com/2006/08/09/technology/09aol.html?pagewanted=all (Accessed 2014, January 6th)
2. Cavoukian, A.: Privacy by Design – The 7 Foundational Principles: http://privacybydesign.ca, (Accessed 2014, January 6th) (2014)
3. Di Vimercati, D. C., S., Foresti, S., Livraga G., Samarati, P.: Data Privacy: Definitions and Techniques. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. 20:6, 793-817 (2012)
4. Flanagan, Howe, Nissenbaum: Enbodying Values in Technology: Theory and Practice. In(eds.): J. V. Hoven & J. Weckert: Information Technology and moral Philosophy. Cambridge University Press, pp. 322 -354 (2012)

5. Friedman, B. Kahn, P.: Human Values, Ethics and Design. The Human-Computer Interaction Handbook, L. Erlbaum Assoc. Inc., Hillsdale, USA., pp.1177-1201 (2003)
6. Hert. P. D., Kloza, D., Wright, D. et al.: Recommendations for a privacy impact assessment framework for the European Union, PIAF (Privacy Impact Assessment Framework) project, Grant agreement JUST/2010/FRAC/AG/1137 – 30---CE---0377117/00---70, Deliverable D3, November 2012, p.5, http: //www.piafproject.eu/Deliverables.html (Accessed 2014, January 6th) (2012)
7. Koops, B. Cuijpers, C. & Schellekens, M.: D 3.2 Analysis of the legal and Ethical Framework in Open Source Intelligence. Versatile Information Toolkit for End-users Oriented Open Source Exploitation. Project number: FP7 – SEC – GA – 2009 – 242352 (2012)
8. Machanavajjhala, A., Gehrke, J. Kifer, D.: ℓ-Diversity: Privacy Beyond k-Anonymity. ACM Transactions on Knowledge Discovery from Data, Vol. 1, No. 1, Article 3, 1-52 (2007)
9. Nissenbaum, H.: Privacy in Context – Technology, Policy and the Integrity of Social Life. Stanford Law Books, Stanford (2010)
10. Nissenbaum, H.: Values in technical design. In C. Mitcham (ed.), Encyclopedia of Science Technology and Ethics. New York: MacMillam, pp. 66-70 (2005)
11. Nissenbaum, H.: How computer systems embody values. Computer – innovative technology for professions, March 2001, pp. 117-119 (2003)
12. Raguse, M., Meints, M., Langfeldt, O., Peissl, W.: Prepatory Action on the enhancement of the European industrial potential in the field of Security research.Tech. report, PRISE (2008) (Accessed, 2014, January, 6th), http://www.prise.oeaw.ac.at/publications.htm
13. Sweeney, L.: K-anonymity: a model for protecting privacy. International journal of uncertainty fuzziness knowledge based systems 10, vol. 5, 557-570 (2002)
14. Tavani, H.: Informational privacy, data mining, and the internet. Ethics and Information Technology. 1, 137-145 (1999)
15. Thuraisingham, B. : ACM SIGKDD Explorations Newsletter, vol. 4, issue 2, December 2002, 1-5 (2002)
16. Van Den Hoven, J.: ICT and Value Sensitive Design. *IFIP International Federation for Information Processing,* Vol. 233, The Information Society: Innovations, Legitimacy, Ethics and Democracy, eds. P. Goujon, Lavelle, S., Duquenoy, P., K. Laurent, V. Boston Springer (2007)