APF 2014 - Opinion papers

Positions & views on different aspects of the proposed EU data protection reform package

Final, June 2014

European Union Agency for Network and Information Security www.enisa.europa.eu

Forward

In January 2012 the European Commission proposed a reform of the EU's data protection legislation. The reform consists of a draft Regulation setting out a general EU framework for data protection and a draft Directive in the police and criminal justice sector. In October 2013 the European Parliament backed the Commission's proposals and in March 2014 it gave its final positive vote on the reform.

Following the aforementioned developments, ENISA invited all interested EU Data Protection Authorities (DPAs) to present relevant opinion papers in its 2014 Annual Privacy Forum (APF) event. The DPAs wehere in particular asked to express their views/opinions on certain elemens of the proposed Regulation that they deem critical, based on their experience and everyday practice in the field.

To this end, a specific panel was organised within APF, where four DPAs, namely the DPAs of Belgium, Spain, Greece and the Czech Republic, presented their opinion papers. The topics addressed were anonymization/pseydonimization, personal data breach notifications, security and privacy by design. The full opinion papers of the panel's participants are included in this special APF 2014 edition.

Table of Contents

Forward ii

1 Position paper from the Belgian Privacy Commission ("the BPC") on the concept of pseudonymous data as introduced by the European Parliament legislative resolution of 12 March 2014 on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data General Data Protection Regulation), Romain ROBERT, legal adviser, Belgian Privacy Commission 1

2 Personal data breach notification: some lessons learnt pending the new data protection regulation, Manuel García Sánchez, Departamento Internacional , Agencia Española de Protección de Datos 7

3 Data Loss Prevention Systems: Security vs Privacy, Konstantinos Limniotis and Georgia Panagopoulou, Hellenic Data Protection Authority 11

4 ORG Information System in the System of Basic Registers - Czech Republic, Eva Vrbová, Czech Data Protection Authorit 14

1 Position paper from the Belgian Privacy Commission ("the BPC") on the concept of pseudonymous data as introduced by the European Parliament legislative resolution of 12 March 2014 on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data General Data Protection Regulation)

Romain ROBERT, legal adviser, Belgian Privacy Commission

This paper intends to develop the point of view of the Belgian Privacy Commission (BPC) regarding the introduction of pseudonymous data, as explained in its Opinion 10/2014 of 5 February 2014 on the General Data Protection Regulation proposal as adopted by the LIBE Committee and the European Parliament[1]

A. Starting point: the concept of personal data

The concept of personal data is a key concept of Directive 95/46/EC[2], since it delimits its scope. According to the Directive (that is still the legal text in force), this concept is to be understood 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".[3]

Over the years the concept of personal data has been discussed and was interpreted in various ways.[4] The notion is indeed not only complex to define in legal terms[5], but it also requires further clarification due to the evolution of data processing in the digital environment.[6] One of the questions raised was whether data which were not directly linked to an individual's civil identity (name, surname) could still be qualified as personal data. In other words, should one consider that an individual is still identifiable within the meaning of the Directive when there is no information relating to his/her civil identity ?

---

[1] Opinion on the draft regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data, as voted by the LIBE Committee of the European Parliament on 17 October 2013; http://www.privacycommission.be/sites/privacycommission/files/documents/Opinion%2010-2014.pdf

[2] Article 2 (a) of 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.

[3] Article 2 (a) of the Directive should be read together with Recital 26: "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".

[4] See for example EU Study – Legal analysis of a Single Market for the Information Society, New rules for a new age ?, DLA Piper, pp. 18-22.

[5] For examples of different implementations of the concept of personal data in some Member States, see Mario Viola de AZEVEDO CUNHA, "Review of the Data Protection Directive: is There a Need (and Room) For a New Concept of Personal Data ?", in *European Data Protection: in Good Health,* Springer, 2012, p. 267.

[6] See Article 29 Working Party, Opinion 4/2007 on the concept of personal data, WP 136, 20 June 2007; EUCJ, *Scarlet vs. Sabam,* C-70/10, 24 November 2011, recital 51, stating that IP addresses "are protected personal data because they allow those users to be precisely identified".

An answer to such question was given by the Article 29 Working Party, which stated in its Opinion 4/2007 that "in general terms, a natural person can be considered as "identified" when, within a group of persons, he or she is "distinguished" from all other members of the group". Therefore, although a person's name is often the most common identifier,[7] identifying individuals no longer necessarily requires the disclosure of their identity in the narrow sense (e.g. through their name).[8]

B. Anonymous data: a dead end?

According to Recital 26 of Directive 95/46, "the principle of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable". Therefore, the Directive is not applicable to anonymous data since they are not personal data. To be considered as anonymised, a data must therefore be processed in such a way that the data subject can no longer be identified.

The Directive does not clarify how such a de-identification process should or could be performed. In this respect, Recital 26 states that "codes of conduct within the 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".

International standards such as ISO 29100 define anonymisation as the "process by which personally identifiable information (PII) is irreversibly altered in such a way that a PII principal can no longer be identified directly or indirectly, either by the PII controller alone or in collaboration with any other party".[9]

The Article 29 Working Party recently issued an opinion on Anonymisation Techniques[10], in which it concluded that "the underlying rationale is that the outcome of anonymisation as a technique applied to personal data should be, in the current state of technology, as permanent as erasure, i.e. making it impossible to process personal data".

On the basis of the abovementioned definition of personal data, the absence of commonly used identifiers (e.g. names, civil registry numbers, ID number,…) cannot in itself guarantee that the dataset is anonymous and that re-identifability will not be possible: anonymous data must not allow anyone to distinguish ("single out") a person from another.[11]

According to Recital 26 of the Directive, "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". Therefore, the more the terms "means likely reasonably to be used" are interpreted as requiring minimum efforts, the larger the scope of the Data Protection Directive –and the future General Data Protection Regulation (GDPR)- will be.

---

[7] Opinion 4/2007 of WP 29, p. 13.
[8] Report on the application of data protection principles to the worldwide telecommunication networks, by Prof. Yves POULLET and his team, for the Council of Europe's T-PD Committee, point 2.3.1, T-PD (2004) 04 final; EUCJ, *Lindqvist*, C-101/2001, 06 November 2003, §27.
[9] http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html
[10] Opinion 05/2014 on Anonymisation techniques adopted on 10 April 2014, WP216.
[11] Additional EDPS comments on the data protection reform package, § 5; https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Comments/2013/13-03-15_Comments_dp_package_EN.pdf

Clearly, such considerations may give rise to some difficulties in practice, since they imply that "complete anonymity" can be difficult to reach, as shown in several cases and research papers.[12] For example, a recent study showed that re-identification of a large dataset of mobile location data was possible even after the data were –allegedly- "anonymised".[13] Therefore, the line between complete anonymity and potential "re-identifiability" is sometimes difficult to draw.

C. Introduction of the pitfall concept of pseudonymous data by the GDPR

The GDPR Commission's proposal[14] defines a personal data as "any information relating to a data subject" being "an identified natural person or a natural person who can be identified, directly or indirectly, by means reasonably likely to be used by the controller or by any other natural or legal person, in particular by reference to an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person". [15] As a result, there is no big change in the definition of personal data, except that a clarification of the "reasonableness" test has been introduced in the definition of personal data.

a. The new concept of pseudonymous data

The compromise amendments voted by the LIBE Committee[16] and adopted by the European Parliament (hereafter 'EP') on 12 March 2014 propose a slightly different definition of personal data.[17] More importantly, the text adopted by the EP introduces the new concept of pseudonymous data. These are defined as "personal data that cannot be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organizational measures to ensure non-attribution".[18]

The EP text confirms that pseudonymous data are personal data, and that they are therefore part of the scope of the Regulation.[19] Here again, one can see here that the definition of identity is at stake, since identity can no longer exclusively be defined in connection with a name, a surname or an

[12] Latanya SWEENEY, "Foundations of Privacy Protection from a Computer Science Perspective" Proceedings, Joint Statistical Meeting, AAAS, Indianapolis, IN. 2000; Paul OHM, "Broken Promises of Privacy: Responding to the Surprising failure of anonymisation", 57 UCLA Law Review, 2010, 1701 (in particular pp. 1716-1730); see also the examples of re-identification cited by the Working Group 29 in its Opinion on Anonymisation Techniques here above mentioned.
[13] Yves-Alexandre de Montjoie, César A. HIDLAGO, Michel VERLEYSEN, Vincent D. BLONDEL, "Unique in the Crowd : The privacy bounds of human mobility", Nature srep. 3, 1376; DOI:10.1038/srep01376 (2013).
[14] European Commission proposal of 25 January 2012, for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), COM (2012) 11 final.
[15] Article 4.1 and 4.2 of the GDPR.
[16] (COM(2012)0011 – C7 0025/2012 – 2012/0011(COD))
[17] According to article 4(2), "personal data" means any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, unique identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social or gender identity of that person;
[18] Recital 24 of the GDPR as voted by the EP, states that "When using online services, individuals may be associated with online identifiers provided by their devices, applications, tools and protocols, such as Internet Protocol addresses or cookie identifiers. This may leave traces which, combined with unique identifiers and other information received by the servers, may be used to create profiles of the individuals and identify them. 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."
[19] See additional EDPS comments on the data protection reform package.

address[20], especially in a digital world, where our digital identity can define us better than our civil identity.[21] Although the identifiers commonly used to distinguish individuals, such as name, surname and address (allowing for a link with "civil identity") are replaced by a pseudonym that does not immediately reveal an individual's "civil identity" (or "real-world" identity[22]), pseudonymous data make it possible to single out individuals and to treat them differently. Therefore, the mere use of pseudonymous data does not guarantee that there will be less intrusion into one's privacy or that the processing of (pseudonymous) personal data is potentially less harmful.

Moreover, the introduction of a new concept in the GDPR such as pseudonymous data can be confusing since it defines a subtype of personal data and might add to the existing problem in distinguishing personal from non-personal data, as already explained here above.

For these reasons, the BPC is against the introduction of the new concept of pseudonymous data, which could be a Trojan horse in the Regulation, leading to the non-application of several of its provisions.

b. <u>The "light regime" proposed for pseudonymous data</u>

Since pseudonymous data remain personal data, the BPC advises against any regime that would exempt the processing of such data from the data protection principles. Such exemptions were proposed in various amendments of the LIBE Committee.

The text adopted by the EP does not include all these amendments. However, Recital 58a states that "Profiling based solely on the processing of pseudonymous data should be presumed not to significantly affect the interests, rights or freedoms of the data subject". Moreover, Recital 38 of the GDPR as voted by the EP provides that pseudonymous data processing operations should *ipso facto* be "presumed to meet the reasonable expectations of the data subjects", which could lead to the elimination of the balancing of represented interests under Article 6.1.f of the GDPR. Finally, article 10 of the GDPR modulates the right of access when it comes to pseudonymous data, on the basis of a flawed reasoning. [23]

Data pseudonymisation should only be considered a technical measure that can be taken into account when balancing represented interests. According to the BPC, digital identity must benefit from the same protection as civil or traditional identity. It is essential that these processing

[20] Report on the lacunae of the Convention for the protection of individuals with regard to automatic processing of personal data (ETS No 108) resulting from technological developments, by Jean-Marc Dinant, Cécile de Terwangne, Jean-Marc Moiny, Yves Poullet, Jean-Marc Van Gyzeghem, p. 20.
[21] However, confusion seems still to exist as the new Recital in the text of the GDPR as adopted by the EP states that "Where profiling, whether based on a single source of pseudonymous data or on the aggregation of pseudonymous data from different sources, permits the controller to attribute pseudonymous data to a specific data subject, the processed data should no longer be considered to be pseudonymous".

[22] Proposed draft EU General Data Protection Regulation and "law enforcement" Directive, Comparative analysis of the ICO of the European Commission text and the European Parliament's LIBE Committee amendments, see comments under the definition of pseudonymous data; see also ICO's definition of Pseudonymisation in Anonymisation: managing data protection risk code of practice, defining pseudonymisation as "The process of distinguishing individuals in a dataset by using a unique identifier which does not reveal their 'real world' identity."
[23] See also Article 10 of the GDPR text as voted by the EP, stating that "If the data processed by a controller do not permit the controller or processor to directly or indirectly identify a natural person, or consist only of pseudonymous data, the controller shall not process or acquire additional information in order to identify the data subject for the sole purpose of complying with any provision of this Regulation." The BPC notes that giving the right of access will not necessarily be more difficult or impossible for the sole reason that the data subject does not use it "real world" identity with his/her relations with the data controller.

operations are fully subject to data protection law, which safeguards the balancing of represented interests and transparency in respect of individuals. Undeniably, profiling based on pseudonymous data could adversely affect an individual's interests, rights and freedoms, and its access to a service or content. In the same way, decisions taken on the sole basis of a unique identifier not being the name or surname of the data subject can be discriminatory even if only pseudonymous data are processed.

For these reasons, the BPC is of the opinion that the distinction between pseudonymised data and other personal data is not appropriate and might compromise the protection guaranteed by the GDPR.

c. Key-coded data used in historical, statistical or scientific historical, statistical or scientific research

Pseudonymised data may also be used in the context of historical, statistical or scientific research, where the purpose of the processing is to increase global knowledge and not to treat individuals differently. The use of pseudonymised/key-coded data for these activities is already subject to specific principles under Belgian law.[24]

The use of pseudonymous data for historical, statistical and scientific research is also addressed in the text adopted by the EP. The relevant provisions stipulate that when it is not possible to perform the research with anonymous data, pseudonymous (or key-coded) data may be used if all necessary measures are taken to prevent unwarranted re-identification of the data subjects.[25] In this specific context, the BPC is of the opinion that pseudonymisation -considered as a minimization technique- should be encouraged.

d. Pseudonymisation should be encouraged as a privacy-enhancing technique, in the context of privacy by design and in order to secure access to the data

While the concept of pseudonymous data should not be accepted when it provides for a "light regime" under the GDPR, the process of pseudonymisation can actually be an efficient way to respect some of the GDPR principles such as privacy by design, security, data minimization, and proportionality. Therefore, the term "pseudonymisation" should be preferred over the notion of pseudonymous data.

For example, article 33.3 e) of the GDPR mentions the process of pseudonymisation as a manner to ensure security of the data: while pseudonymisation is not a method of anonymisation, since it merely reduces the linkability of a dataset with the original identity of a data subject, it can be a useful security measure for the same reasons.[26] Pseudonymisation can also be a way to ensure that data affected by a breach are adequately protected against identification in the context of data breach provisions (articles 31 and 32 of the GDPR). Besides, pseudonymisation could be considered

[24] See Belgian Royal Decree of 13/02/2001, Chapter II.
[25] Article 81.2 a of the EP text adopted on the GDPR.
[26] Opinion of the Article 29 Working Group, p. 3.

as a best practice in order to comply with the provisions relating to privacy by design (article 23 of the GDPR).[27]

As a conclusion, the BPC is against the introduction of the new notion of pseudonymous data into the regulation since this would blur the distinction between personal and anonymous data and would weaken the protection granted to individuals, especially if exemptions for pseudonymous date are voted. However, the BPC encourages the use of pseudonymisation techniques in the Regulation as a means to meet some of the data protection principles, such as data security, data minimization, privacy by design or proportionality.[28]

[27] Likewise, Recital 122a of text of the EP resolution states that "a professional who processes personal data concerning health should receive, if possible, anonymised or pseudonymised data, leaving the knowledge of the identity only to the General Practitioner or to the Specialist who has requested such data processing."
[28] Article 29 Working Party, Document 1/2009 on pre-trial discovery for cross-border civil litigation. This document encourages the controller to restrict disclosure, if possible, to anonymised or at least pseudonymised data as a first step.

2 Personal data breach notification: some lessons learnt pending the new data protection regulation

Manuel García Sánchez, Departamento Internacional , Agencia Española de Protección de Datos [1]

Mandatory notification of personal data breaches was included in the 2009 reform of the E-privacy Directive[2] being the first time an obligation of that kind was introduced at European level. Even though originally this requirement was limited to providers of publicly available electronic communications services, the foreseen extension to a general obligation included in the proposal for a Data Protection Regulation[3] makes it interesting to have an insight on the developments that have taken place since then.

The Directive

Directive 2002/58 imposes providers the obligation to notify personal data breaches to national competent authorities as well as to subscribers or individuals insofar as they can likely be adversely affected by the breach. The scope of the obligation is broadly defined[4] since almost every security incident affecting personal data qualifies for notification to the competent authority while there is no rule defining thresholds for notification. In the same line, the reference to likely adverse effects to subscribers or individuals leaves wide margin of appreciation for both the obliged entity and the competent authority.

Similar margin for appreciation results from the general reference to national competent authorities. Since the obligation was created to deal with security incidents affecting personal data, it might have seemed more logical to explicitly establish data protection authorities as the competent ones; eventually the legislator decided to leave in the hands of the Member States the final decision. Competent authorities have been charged with the obligation of providing guidance and instructions on the format and procedures related to the manner in which notifications need to be done as well as to carry out audits to ensure effective compliance and to impose sanctions in the event of an infringement of the law. Last but not least, telecommunications providers are obliged to maintain an inventory of data breaches comprising the facts, effects and remedial action taken for each data breach that needs to be at the disposal of the national competent authority.

The legislator's intention can be interpreted in the sense of setting up a system covering the widest range of security incidents affecting personal data while restricting to some degree notifications to individuals. The role of the competent authority is defined in a multidimensional way since they are entrusted with a complete set of tasks covering from guidance and support to the monitoring and checking of compliance. In that sense, the absence of notification thresholds may be interpreted in the sense that a complete set of notifications will be offering a full picture of the situation easing the advisory role of the authorities.

[1] The opinions expressed in this paper are those of the author

[2] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:2002L0058:20091219:EN:HTML

[3] A similar initiative has been included in the draft proposal for a Directive Proposal for a Directive of the European Parliament and the Council on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and the free movement of such data.

[4] 'personal data breach' means a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to, personal data transmitted, stored or otherwise processed in connection with the provision of a publicly available electronic communications service in the Community

Certainly striking is the fact that, as opposed to what was done in other cases, the obligation was limited to the providers of publicly available electronic communications services, putting out of reach important sectors like providers of information society services[5].

Implementing measures and coordination efforts

With the aim to ensure consistency in the implementation of this obligation and to avoid legal uncertainty and excessive implementation costs, the European Commission issued in 2013 a Regulation[6] aiming to define common parameters for notifications to both national competent authorities and individuals or subscribers. The Regulation also included a first approach on the cases in which data might be considered unintelligible for the purpose of avoiding notification to individuals and, more important, setting out the minimum set of information to be included in both types of notification.

However, the content of both notification models was vaguely defined hampering the possibility of taking advantage of having a basic common form shared by all competent authorities. This could also impact the expectations of getting consolidated statistics related to the total amount, distribution of notifications and relevant features of the breaches at both Member State at EU level that could be certainly of great help when assessing the effectiveness and impact of the measure on a better protection of individual rights.

Since 2009, the process of transposal and application of the Directive has been accompanied by significant attempts to coordinate efforts by relevant stakeholders like the Article 29 Working Party[7], the European Commission and other stakeholders[8]. Besides that, the implementation process at the national level has also showed specific shortcomings, as explained below.

National competent authorities

Reference has been made to the fact that the Directive leaves room to the Member States to select the national competent authority. In that sense, some Member States decided to appoint an authority different than the Data Protection Authority that, in principle, seems to be the most convenient option for reasons like the specific connection with the competences conferred by the data protection law as well as the specific know-how derived from their daily work. In these cases, the division of competences between the competent authority on notification and the data protection authority also involves the need to set up collaboration mechanisms at national level as well as other tools at EU level since non-DPA competent authorities are not members of bodies like the Article 29 Working Party.

The Directive also charges competent authorities with a wide set of competences ranging from advising to enforcing compliance. This, in absence of proper resources, might prove to be difficult provided that some competent authorities may not have sufficient technical and human resources to cope with the task. As a result, the exercise of the conferred powers can present asymmetries at national level affecting the level of protection of individuals as well as the real possibilities to ensure equal treatment for the providers irrespectively of their location. Providers can be also affected by different compliance requirements at national level – different forms and notifications methods as well as slightly different transpositions – placing an extra burden in cases like providers offering services in several Member States.

[5] Despite of the request that was made in that sense by the European Parliament during the negotiations.
[6] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:173:0002:0008:en:PDF
[7] http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2011/wp184_en.pdf and http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2012/wp197_en.pdf
[8] For ENISA: http://www.enisa.europa.eu/activities/identity-and-trust/risks-and-data-breaches/dbn

Notification thresholds

Even though the existing figures are not meeting previous expectations about competent authorities "flooded" by notifications, the decision of not setting up notification thresholds is in itself a risk for the proper management of the notification system. In that sense, the possible loss of useful information resulting from the establishment of notification thresholds could be compensated by reinforcing the role of the inventory within the system. Thus, a two-level notification system could comprise the obligation for the provider of maintaining a detailed record of each personal data breach – the inventory – while limiting the direct notification to the competent authority to those breaches above a given threshold. That said, the establishment of common thresholds should be the result of a meticulous selection of criteria subject to periodical review.

Severity assessment criteria

Severity assessment criteria have been identified from the beginning as a critical issue for both competent authorities and providers but to this day this is still an ongoing issue. Despite of the efforts made by some stakeholders, there is as yet no widely accepted methodology[9] allowing authorities and providers to properly assess the severity of the breach and the need for notifying individuals.

A common methodology will ensure that similar breaches would get the same assessment irrespectively of the location of the provider guaranteeing also the use of harmonized criteria. In any case, there is no doubt that the obligation of establishing assessment criteria lays primarily on the competent authorities as well as the European Commission.

Cross-border data breaches

Cross-border data breaches are also a conflicting issue. No criteria have as yet been defined on how to manage those breaches in terms on who needs to be notified, which competent authority should be assuming the leading role on the investigation and which means should be used for channeling information to all the stakeholders involved. This lack of criteria may result in an increase in the risk of adverse effects for individuals. Coordination efforts are needed between all parties concerned, specifically with regard to information and reporting obligations.

Reluctance to comply

It seems that providers find little incentive in properly notifying insofar as they might be in some cases more worried about possible reputational costs derived from notifications than to the possible impact on individuals. It could be said that in some cases providers are living in a permanent "state of denial" with regard to data breaches resulting in a very limited number of notifications.

Even though there is still no empirical evidence, it seems that only a fraction of the personal data breaches affecting providers are duly notified. It is clear that some providers need to understand that they will in any case experience personal data breaches so they need to prepare themselves on how to protect personal data and to properly face data breaches in order to meet their obligations to their subscribers.

In the same line, additional efforts should be made by competent authorities to ensure that not only all notifications are done in time and according to legal requirements in place but also to ensure that all stakeholders involved are well aware of the benefits provided in terms of protection of individuals and prevention and management of future risks.

[9] http://www.enisa.europa.eu/activities/identity-and-trust/library/deliverables/dbn-severity and http://www.cnil.fr/fileadmin/documents/en/CNIL-ManagingPrivacyRisks-Methodology.pdf

To conclude

There is no doubt that the initiative of including the obligation to notify data breaches for providers of publicly available telecommunications services was a sound decision irrespectively of the limited scope of application. However, it is already possible to identify a number of shortcomings in its practical implementation that need to be properly addressed considering also the foreseen extension of its scope to cover all relevant sectors.

In that sense, it would be desirable giving national data protection authorities exclusive competence on this field while ensuring proper coordination and harmonization in terms of technical tools and procedures at EU level. This would require agreeing on common formats for notification based on internationally recognized technical standards allowing secure data exchange and compilation for statistics purposes.

Defining common thresholds and procedures for managing local and cross-border notifications as well as common severity assessment methodologies to be used for both providers and competent authorities would also result in great benefits for the individuals.

And last but not least, awareness raising efforts combined with effective guidance and compliance may lead in favor of significant reduction of the impact of personal data breaches across the EU. All in the understanding that a data breach notification system should be just an element of an integrated, security oriented approach guaranteeing adequate protection of individuals' fundamental rights to privacy and data protection

3 Data Loss Prevention Systems: Security vs Privacy

Konstantinos Limniotis and Georgia Panagopoulou, Hellenic Data Protection Authority

Abstract. Data loss prevention systems, from a data protection point of view, are studied in this paper. More precisely, data protection issues that arise from the use of data loss prevention systems as a security control are being considered, stating appropriate measures to be taken in order to apply the privacy-by-design principle in the operation of such systems.

Introduction

It is widely known that organisations should put much effort on ensuring the compliance with the personal data protection legislation; apart from the aforementioned obligation, avoiding data breaches is of high importance as the business impact of a data breach could be disastrous. To this goal, several organizational and technical measures are being implemented, whose effectiveness should be constantly evaluated.

Among other security controls, data loss prevention (DLP) systems have a prominent role. By contrast to intrusion detection systems, which aim at scanning incoming traffic, DLPs focus on internal traffic and outgoing data - i.e data that leave the company. Hence, by this way, the company prevents any potential unauthorized use or transmission of proprietary data, independently from whether such an action is unintentional or not. It should be pointed out that the Hellenic Data Protection Authority has received some inquiries, as well as notifications, from data controllers regarding the use of DLPs, mainly as a response to recommendations regarding the need for applying high-level security measures.

However, the special nature of DLP systems poses several risks with respect to the privacy of employees (and, probably, of third parties); this stems from the fact that such systems automatically gather large amounts of personal data and, thus, investigating such data may infringe the privacy. Hence, the adoption of a DLP solution is a decision, which should be taken after careful assessment with all involved stakeholders.

This paper focuses on data protection issues that arise from the use of DLP systems that monitor the outgoing e-mail traffic. More precisely, the aim of the paper is to exhibit the privacy risks that occur when such a DLP system is in place, as well as to propose general rules to mitigate those risks, according to the Privacy-by-Design principle which is a crucial factor for data protection (and is further strengthened in the proposal for the new General Data Protection Regulation [4]}).

The paper is organized as follows; In Section 2 a short overview of the DLP technology is given, describing the main content-analysis techniques that are being met, while Section 3 emphasizes on the effects to the employees' privacy, providing guidelines for a proper use of such systems. Finally, concluding remarks are given in Section 4.

Data Loss Prevention Systems: A short overview

In general, a DLP solution may monitor data in motion (e.g. outgoing e-mails), data at rest (e.g. when users save data on network folders) and data in use (e.g. as users access files) [2]. In any case, a content-analysis is being performed, towards deciding whether a specific action is admissible or not

(according to well-defined policies). The widely used content-analysis techniques include (see, e.g. [5]):

1. Keyword matching, which is based on scanning the content for specific kewords from a list.
2. Rule based and regular expression matching, which is the most common technique. It is based on analyzing the content for specific rules — such as 16-digit credit card numbers.
3. Database fingerprinting, which takes either a database dump or live data from a database and seeks for exact matches.
4. Machine learning (or other statistical-based) algorithms, to analyze the content via an appropriate classifier which has been appropriately trained (similarly to the spam filters).

The core of any DLP solution is the network monitoring, whereas email integration is also a main component which strives to protect from data breaches via electronic mails. However, an automated scanning of outgoing e-mails, with further investigation of those that have been characterized (or intercepted) as suspicious, increase the danger of employees' privacy, since e-mails should benefit from the same protection of fundamental rights as traditional paper mail ([1]). Hence, there is a trade-off between security (i.e. customer's personal data protection) and employees privacy (employees personal data protection), which should be appropriately treated.

Protecting Privacy

In this Section, we describe a set of rules that have to be applied when a DLP solution that is based on outgoing e-mail scanning is under consideration. Our approach is based on forcing the well-known data protection principles (such as proportionality) to such systems - which is certainly a nontrivial task. The goal is to apply the Privacy-by-Design approach in the adoption of a DLP solution, characterized by proactive rather than reactive measures ensuring privacy [3].

More precisely, such a processing will be legitimate if it is necessary for the purpose of the legitimate interests pursued by the companies (data controllers), provided that the processing does not violate the rights and freedoms of the data subjects. To this end, the following steps seem to be prerequisite:

- Before implementing a DLP solution, a Privacy Impact Assessment (PIA) should be conducted, to assess the privacy and data protection impacts of such a choice (with the view of examining all possible alternatives to prevent those impacts). The outcome of the PIA should be a fully justified decision regarding the necessity or not of the DLP solution.
- Ensuring transparency of the system is essential. The employees should be provided with full information regarding the operation of the DLP system, whereas accurate internal rules and regulations should be in place. Clearly, the corresponding policy should explicitly refer to the fact that the work e-mail account is intended for work purposes and not for personal communications, so as to avoid monitoring personal communication; hence, the company may scan only e-mails stemming from e-mail addresses which are given by itself (and not e-mails that are exchanged via personal accounts - e.g. via webmail).
- It is crucial to choose appropriate techniques for identifying suspicious e-mails, so as to minimize false positives alarms.
- It is important to force appropriate access rights so as to ensure that only authorized trustworthy persons may have access to data stored by DLPs. Clearly, the need-to-know principle should be applied (for instance, there is no need for a system administrator to have access to these data). To this goal, it should be also pointed out that personal data in the warnings (alarms) should be minimized.

- The DLP log data may, in most cases, contain personally identifiable information, and is therefore subject to the personal data protection legislation, therefore employees must be able to exercise their subject's rights to information, access, deletion to their personal data contained in DLPs log files.
- Third parties' personal data should not be processed by the DLP system. Note that an e-mail may contain such data (for instance, in case of e-mail conversations).
- Data that have been investigated and considered to be innocent, should be immediately deleted.

Note that the above list is not exhaustive; a proper decision of whether a DLP solution of such type is good choice or not should be made, in general, on an ad-hoc basis, according to the aforementioned PIA outcome. The PIA should also determine the appropriate measures that accompany the deployment of the solution.

Conclusions

DLP systems can be an important security control for securing personal data processing but in the same time poses several risks with respect to the processing of employees and third parties personal data. Applying appropriate measures is prerequisite for mitigating those risks and striking the proper balance between legitimate interests of organizations to protect their data and the fundamental right to the protection of individuals' personal data.

Acknowledgements

The authors would like to thank the anonymous reviewers for their comments and suggestions, which helped to improve the presentation of the manuscript.

References

1. Article 29 - Data Protection Working Party: Working document on the surveillance of electronic communications in the workplace. May 2002 (https://www.enisa.europa.eu/activities/cert/support/proactive-detection/proactive-detection-report).
2. Daley, M. J., Fey, L. C., and Fashing, D.: Exploring Data Loss Prevention Systems for Legal Holds and e-Discovery. Information Management, vol. 44. n. 5, pp. 26--30, Sept. 2010, ARMA International.
3. Enisa Report: Proactive Detection of Network Security Incidents. Deliverable - 2011-12-07 (http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2002/wp55_en.pdf).
4. European Commission: Proposal for the EU General Data Protection Regulation. (http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf)
5. Rick, M.: Understanding and selecting a data loss prevention solution. Securosis Report, 2010.(https://securosis.com/assets/library/reports/Understanding_and_Selecting_DLP.V2_.Final_.pdf)

4 ORG Information System in the System of Basic Registers - Czech Republic

Eva Vrbová, Czech Data Protection Authority

1. Introduction

Identification of citizens in the Czech Republic:
The basic identifier of every citizen of the Czech Republic is the **birth number**. It is a unique parameter which can reveal information like date of birth, sex, and in some cases even place of birth. This makes them easily vulnerable to misuse. It was firstly introduced in 1953, in paper form then, for communication with public authorities. The birth number, however, contains elements that can be qualified as personal data.

With the expansion of ICT and electronic databases, the birth number was more and more used in all cases where individual identification was requested. This number, given its nature (unequivocal for each individual) has been used as key item in the majority of databases operated by public as well as private institutions. This enabled to match different databases just through the birth number. Uncontrolled use of birth numbers and database interconnection raised questions in the field of personal data protection and identity theft.

As the accessibility of ICT grew and database systems improved, public authorities gathered more and more electronic data. The public administration has created a great amount of databases. Each public administration segment had its own set of databases in which same items often repeated (address, marital status, spouse, children) that in many cases varied in terms of content. Creation of data networks and interconnection of computers brought about growing error rate of these duplicities. Rectification of these duplicities and inconsistencies in all databases is a time-consuming exercise, demanding also administratively. Further, if any data gets rectified in one database, the parameter has to be updated in other relevant databases, too. Additionally, the maintenance of these databases is technically and financially very demanding which leads us to the conclusion that the system of data management in public sector calls for change and a new system eliminating these problems will have to be designed.

The effort to unify the content and format of data, to remove duplicities, as well as to prevent them has lead to the development of the basic registers system. It was called for a system that would

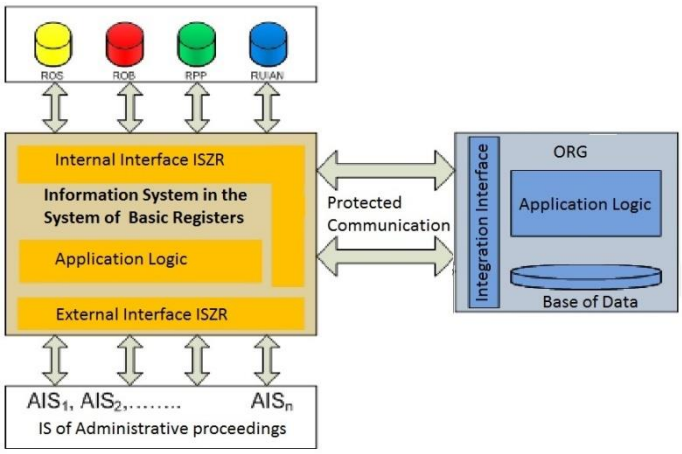
disperse the interconnection of personal data and reduce their relatively easy misuse through multiple identity of a citizen. Basic registers have become a system like this. They are a secure and up-to-date database on citizens and public as well as non-public entities.
Essential move to create and put into operation such a system was the adoption of the Act No. 111/2009 Coll., on basic registers and of the Act No. 227/2009 Coll. early in 2009.
Implementation of the basic registers project is funded from the EU structural funds.

What does the system offer?
Public administration basic registers are one of the fundamental pillars of the modern eGovernment in the Czech Republic, i.e. of the process of public administration computerization. Citizens and entities have accepted the basic registers system as a state-of-the-art component of administrative management. The objective is to render the administration more effective and to provide accessibility from almost everywhere and anytime. The systems must, at the same time, meet criteria for effective, secure, and transparent exchange of the so-called reference data that must be accurate and up-to-date.
Apart from the enhanced effectiveness, where officers do not need to verify whether the data are up-to-date and correct, the proceedings sped up whilst the administrative burden declined. Citizens spend less time at the counter. The state works more a more effectively, you save time and money.
Basic registers are secure and up-to-date database on citizens and public and private entities. Implementation of these registers will reduce fragmentation, disunity and multiple use of personal data in basic registers of public administration.

There are four basic registers surrounded with other system components which are separated from each other in terms of hardware and location and are administered by different institutions.

- Population register (Interior Ministry is the administrator) – relevant reference data on the Czech citizens, foreigners with residence permit, or foreign property owners
- Rights and obligations register (Interior Ministry is the administrator) – data on competence of public administration bodies and on the rights and duties of individuals
- Register of persons (Czech Statistical Office is the administrator) – data on legal persons, physical persons doing business, or public administration bodies
- Territorial identification, addresses and property register (Czech Geodetic and Cadastral Office is the administrator) – data on basic territorial elements, e.g. state territories, districts, municipalities or parts of municipalities, plots of land, streets
- Basic registers information system (Interior Ministry shall be the administrator) – four basic registers are in operation within its framework. Currently administered is by the Basic Registers Administration.
- Convertor of physical persons identifiers (Office for Personal Data Protection is the administrator) – key project from the data privacy viewpoint. Thank to the convertor, it will be possible to retrieve information on a citizen practically from every public administration

information system only on the basis of birth number (it will not be possible without knowledge of the birth number)



- ORG communicates expressly and solely with the information system of basic registers (ISZR).

- Basic registers:

- **ROS** – Register of persons.
- **ROB** – Register of population.
- **RPP** – Register of rights and obligations.
- **RUIAN** – Register of territorial identification, addresses and real estate.

It is possible to interconnect the individual registers via the convertor of physical persons identifiers and with assistance of the Basic registers information system.

ORG communicates expressly and solely with the information system of basic registers (ISZR).
Essential element in the basic registers system is the so-called reference data. It is in fact a piece of information taken from the Basic registers system and is used for the relevant administrative proceedings as a guaranteed data, valid and up-to-date that do not need to be verified. Public administration bodies are obliged to use data just from the basic registers and not to require them from citizens.
For public administration institutions basic registers represent updated information on every citizen. Introduction of basic registers removed fragmentation, lack of homogeneity and multiple data storage in basic public administration databases. To citizens, it facilitated significantly communication with public administration bodies as it will be enough to notify any change of the stored data on a single spot only. Also for the public administration this project made their work more transparent and gave the officers more assurance that the data stored are correct.
Basic registers replaced the birth numbers used as universal identifier so far. Still, birth numbers shall remain with the citizens until 2025.
Apart from creating a self-contained system of reference data in the individual registers, also the level of security of our personal data has been enhanced. It means that for example the home address or place and date of birth will be stored on one place only and it will be a reference and valid information. Notifications of data amendments are made daily and afterwards authorized users can download them from their respective systems.
Nowadays, the Basic registers information system is connected to other 2700 information systems of public administration. The entire system of basic registers runs in routine operation 24 hours a day, 7 days a week. It is supposed that the number of connected components will grow in future. Successively, even the commercial and health sectors will receive connection to the basic registers.

2. ORG Information System

The ORG information system forms part of the basic registers. This system shall principally secure the protection of personal data across the basic registers by way of replacing the birth numbers, used so far, with a set of anonymous identifiers. Operation of this system has become new competence of the Czech Office for Personal Data Protection.

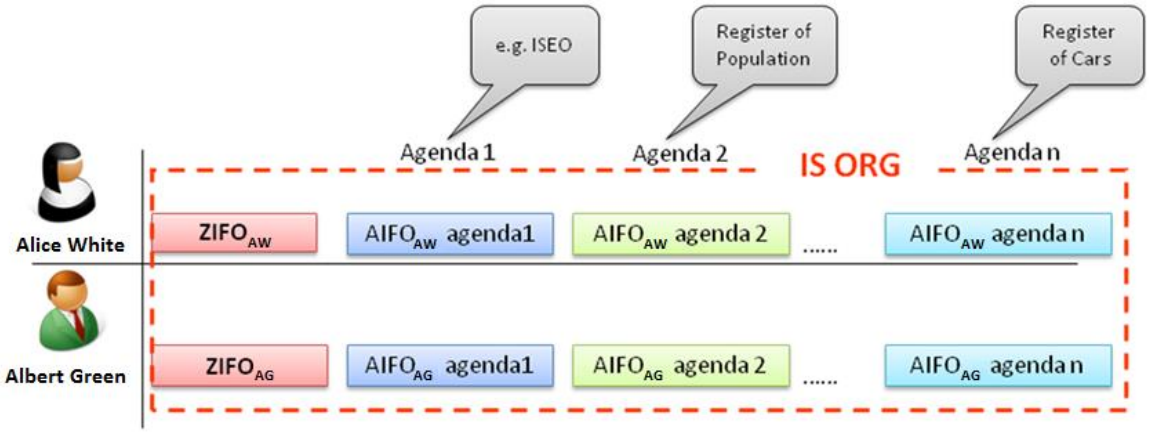
Main sense of the ORG information system is to ensure protection of personal data in the entire basic registers system. This is provided through replacement of birth number (used as physical person´s identifier so far) by a system of non-essential identifiers. These identifiers shall vary for every single type of administrative proceeding or set of proceedings. So that with knowledge of one identifier shall it not be possible to look up physical person´s data in databases related to other proceedings. The single place of storage of all identifiers shall be just the ORG information system. This system however does not contain any other data about physical persons, thus knowledge of all identifiers shall not enable the Office for Personal Data Protection to match them with specific persons.
It is like cutting a photograph to pieces and giving each piece to different persons. Reverse assembly will always be more difficult than to remove it from a photo album.

ZIFO is the source individual identifier. It is a unique non-essential identifier recorded exclusively in the register of source individual identifiers. This identifier cannot be personally obtained it is generated by the ORG system and used for internal purposes of the basic registers system. Nobody shall ask citizens for this “number”, it serves only for internal use within the basic registers.

Safe data sharing within the basic registers is guaranteed through ZIFO and AIFO identifiers as well as owing to the fact that the information system neither contains nor processes any personal data of citizens. A mathematical instrument working in HSM environment has been created to generate ZIFO and AIFO. A security policy has been elaborated including risk analysis.

AIFO is the administrative individual identifier, a unique, likewise non-essential identifier, under which physical persons are registered in a specific agenda information system.
Every Czech Republic´s citizen as well as each foreigner with permanent residence in the Czech Republic is unambiguously identified in the basic registers system through one ZIFO and a group of AIFOs in a way that a self-contained AIFO is generated for each individual agenda (administrative proceeding).
More you can read in [1], [3], [6], [7] .

Example: if anew citizen is born, an Albert Green, the population register asks the Office for Personal Data Protection to generate new ZIFO. This ZIFO is stored in a database. Afterwards, AIFO is calculated and sent back to the population register. There, the AIFO is assigned to the name xxx. At the same time, other institutions (responsible for other administrative proceedings) are informed through the basic registers system about creation of a new item, new AIFO. The Office for Personal Data Protection calculates for each administrative proceeding its own AIFO. At that moment, new citizen Albert Green is introduced in all proceedings necessary, not under his name, but under a number which is different for each proceeding.



1 Security of identifiers

ZIFO is entirely non-public, designed for internal needs of the ORG information system only (for purposes of deriving AIFO and linking AIFO group to the given person). These identifiers are unique for each individual, generated as random chain, not inferred from any personal quality.

AIFO too is a non-public data stored for one thing in the ORG information system with linkage to ZIFO, and for another transferred to AIS. Used internally only, it serves identification of physical persons. Consequently, AIFO has different value for each agenda (administrative proceeding).
Protection of AIFO on the AIS side is then dependent on the method of security for every single AIS.
Security of data sharing within the basic registers is grounded on the principle that:

- ORG contains the identifiers ZIFO and AIFO only;
- ORG neither contains nor processes any citizens´ personal data except the identifiers;
- Online data messages contain only the anonymous non-essential identifier AIFO and the transferred data;
- Matching of data between administrative proceedings and the basic registers is possible via ORG only.

Each agenda (administrative proceeding) or AIS which handles the agenda uses its own group of AIFO. This is completely different to that of another agenda, hence not applicable for personal data matching. Interconnection is possible only through the ORG information system. AIS are involved in the ORG services via an interface of basic registers information system (ISZR) where access authorization is required and checked by way of a rights and authorizations matrix operated for the whole basic registers system by the register of rights and obligations (RPP).

2 Mathematical tools for generating identifiers

A set of mathematical tools has been designed to generate ZIFO and AIFO. For the sake of security these tools were implemented as application in the HSM operational environment working in the Secure Execution Engine (SEE) mode. Thus, an effect was achieved that ZIFO generating, but also AIFO deriving are possible only in HSM. HSM therefore fulfils the following functions:

- Secure administration and saving of cryptographic keys;
- Provision of quality generator of random numbers for generating keys and ZIFO ;
- Provision of secure environment for operating the mathematical tools.

The mathematical tools enable to scale performance so that the system is trimmed for future bigger demands as to the number of generated identifiers.
The advanced encryption standard is based on [1] and [6].

3 Personal data protection in basic registers

Protection of personal data is provided on several levels:

- Basic registers are separated in terms of data
- ORG information system does not contain any personal data, only a matrix of identifiers
- Messages sent contain only AIFO (non-essential identifier) and factually anonymous data
- User of basic registers or gets only data for which he is authorized

The ORG information system only can match data related to individual administrative proceedings (agenda) through the AIFO matrix. Neither ORG however knows to whom AIFO pertains. Moreover it does not come into contact with personal data at all.

Birth number	Basic identifier of every citizen of the Czech Republic, it contents from 10 (sometimes 9)digits
ZIFO – Source individual identifier - zdrojový identifikátor fyzické osoby	It is a unique non-essential identifier recorded exclusively in the register of source individual identifiers.
AIFO – administrative individual identifier - agendový identifikátor fyzické osoby	Unique, likewise non-essential identifier, under which physical persons are registered in a specific agenda information system.
Act No. **111/2009 Coll.,** and **Act No. 227/2009 Coll**.	Acts about basic registers
ROB: Register of population (Interior Ministry is the administrator) – registr obyvatel	Relevant reference data on the Czech citizens, foreigners with residence permit, or foreign property owners.
RPP: Register of rights and obligations (Interior Ministry is the administrator) – registr práv a povinností	Relevant reference data on the Czech citizens, foreigners with residence permit, or foreign property owners
RUIAN: Register of territorial identification, addresses and property (Czech Geodetic and Cadastral Office is the administrator) – registr osob	Data on basic territorial elements, e.g. state territories, districts, municipalities or parts of municipalities, plots of land, streets
ROS: Register of persons (Czech Statistical Office is the administrator) – registr územní identifikace, adres a nemovitostí	Data on legal persons, physical persons doing business, or public administration bodies
ORG: Convertor of physical persons identifiers information system ORG (Office for Personal Data Protection is the administrator)	Key project from the data privacy viewpoint

ISZR: Basic registers information system (Interior Ministry is the administrator)	Four basic registers are in operation within its framework
AIS: Information system of agenda – agendový informační systém	Information system for each agenda

References:

[1] Vrbová, Eva, Informační systém ORG, conference Egovernment Mikulov, 2011
[2] Vrbová, Eva, Základní registry – 2 měsíce provozu IS ORG, conference Egovernment Mikulov 2012
[3] Vrbová, Eva, IS ORG v systému základních registrů, conference eGovernment Praha, 2012
[4] Vrbová, Eva, IS ORG v systému základních registrů, conference eGovernment Praha, 2013
[5] Vrbová, Eva, Informační bulletin 1/2012 Úřadu pro ochranu osobních údajů - základní registry, document, 2012
[6] Vrbová, Eva, IS ORG information system in the system of the basic registers - document in DSM magazine 2012
[7] Vrbová, Eva, ORG information system in the system of the basic registers - international conference Security and Protection of Information Brno, 2013
[8] Šusta, Antonín, ÚOOÚ jeho nové competence v roce 2010 - Informační systém ORG
[9] Vrbová, Eva, Dopady zavedení IS ORG do života, conference ISSS, Hradec Králové, 2012
[10] Informační system ORG – bezpečnost základních registrů, lecture for ombudsman Brno, 2012

ENISA
European Union Agency for Network and Information Security
Science and Technology Park of Crete (ITE)
Vassilika Vouton, 700 13, Heraklion, Greece

Athens Office
1 Vass. Sofias & Meg. Alexandrou
Marousi 151 24, Athens, Greece

PO Box 1309, 710 01 Heraklion, Greece
Tel: +30 28 14 40 9710
info@enisa.europa.eu
www.enisa.europa.eu