DATA SUBJECT AS AUGMENTED REALITY

Vladimir A. Dokuchaev,
Moscow Technical University of Communications and Informatics, Moscow, Russia,
v.a.dokuchaev@mtuci.ru

Victoria V. Maklachkova,
Moscow Technical University of Communications and Informatics, Moscow, Russia,
v.v.maklachkova@mtuci.ru

Viacheslav Yu. Statyev,
Transport Safety Fund, Moscow, Russia,
svu@rnt.ru

DOI: 10.36724/2664-066X-2020-6-1-11-15

ABSTRACT

This article describes some security problems of the personal data in the context of global digitalization of all life activities of modern state, society and individuals, when the physical object, processes and personalities management are shifted to the management at the level of their digital profile ("digital twin"). Therefore, in view of the transition from physical reality to digital Augmented Reality (AR), the adequacy of the digital profile of a particular person has to be of paramount importance. The recent events of the end of 2019 and the first half of 2020, occurred due to the world COVID-19 pandemic, have forced some states to move away from the standard procedures of processing personal data implied in the General Data Protection Regulation (EU GDPR) for the reason that all the post pandemic circumstances do increase the risk for the data subject. According to the outcomes of the research conducted in the article, along with the model of threats and intruder of personal data security, the authors consider it to be appropriate for information processing personal data systems to develop the model of consequences for the data subject in case of implementation of the actual threats identified in the threat model. On balance, the proposed approach will enable to come to the answer in near future, "Are the digital transformation and digital twins the main risks to natural personal life privacy or not?"

KEYWORDS: *personal data, natural person, data subject GDPR, digital twin, data protection, pseudonymisation, processing, controller, big data, security, privacy, augumented reality, risk, models, system, information quality.*

Information about authors:

Vladimir A. Dokuchaev *(DSc, Prof.), Network Information Technologies and Services, Moscow Technical University of Communications and Informatics (MTUCI), Moscow, Russia*
v.a.dokuchaev@mtuci.ru
Victoria V. Maklachkova, *Moscow Technical University of Communications and Informatics (MTUCI), Moscow, Russia*
Viacheslav Yu. Statyev, *(PhD), Expert Council, Transport Safety Fund, Moscow, Russia*

I. INTRODUCTION

Development of information and communication technologies (ICT), attention and interest in privacy issues and the security of personal data had tremendous influence to all aspects of humanity life. National interests in the information sphere are aimed at ensuring and protecting constitutional human rights and freedoms, as well as the confidentiality of personal data when using ICT/Telecommunications. From this follows the attention paid to the organization of processing and ensuring the security of Personal Data.

'Personal Data' means any information relating to an identified or identifiable natural person ("Data Subject").

An identifiable natural 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, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person [1].

Indirect identification of the data subject data is the basis of the concept of Big Data, which is currently being actively developed in relation to the field of personal data.

The transition to the Digital Economy provides for the creation of a platform for the identification of citizens, within the framework of which it is assumed that there is a digital profile ("Digital Twin") of a natural person.

This raises the question of the need for anonymization, depersonalization of information about an individual. This simplifies the resolution of issues with the processing of personal data, as laws of various countries on the protection of personal data do not consider anonymized information as personal data.

The creation of a "Digital Twin" of a natural person has further increased the severity of the problems that arise when working with personal data.

One of the key points in the field of personal data is the identification of the data subject. Identification of data subject can be carried out on the basis of any form and content of information relating directly or indirectly to a specific or determinable natural person [2].

The events of the end of 2019 - the first half of 2020, which occurred in the world in connection with the COVID-19 pandemic, forced the governments of some countries, for example, Hungary [3], to deviate from the standard procedures for identifying personal data subjects provided for in the General Data Protection Regulation (EU GDPR) [1].

The antithesis of the identification of the data subject is the pseudonymisation of his personal data. So pseudonymisation of personal data in certain legislative acts is defined as "processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person" [1]. In particular, requirements are put forward that personal data being processed must be destroyed or anonymized when the processing goals are achieved or if the need to achieve these goals is lost and personal data is processed for statistical or other research purposes, except for the purposes specified in the relevant legislative acts, when subject to mandatory anonymization of personal data, respectively. It can be stated that the anonymization of personal data and the destruction of personal data are phenomena of the same order.

The process of identifying a natural person as a data subject has a very vague description in regulatory documents of various countries.

In the context of the foregoing, it is necessary to pay special attention to problem that, as the authors think, is relevant and has not been considered in sufficient detail at the present time.

This problem is related with the assessment of harm that may be caused to personal data subjects in case of violation of applicable law. There are no standard recommendations on possible approaches to assessing the harm to data subjects at the present time. Note that this harm is purely individual in the same conditions for different data subjects.

An assessment of the harm that may be caused to the data subject can be done only after determining the consequences (in the terminology of information security) [4, 5] related to the violation of confidentiality, integrity and accessibility of personal data. In essence, this is, first of all, unlawful familiarization and unauthorized distribution of personal data, deterioration in the quality of personal data and untimely access to them. The harm can be different depending on the goals set by one who wants to take advantage of the consequences of violation of the Data Protection Law.

And so, in the context of what has been said, the question can be formulated as follows: "Who identifies the data subject, for what purpose and what is the subject of identification"?

II. PURPOSE OF THE DATA SUBJECT IDENTIFICATION

Identification of the data subject may be handled by the controller (operator of personal data - in Russia) or by a third party who legally or illegally gained access to information about the data subject. The purpose of identifying the data subject on the part of the controller (operator of personal data) is obvious and directly related to the purpose of processing personal data. But the purpose of identifying the data subject on the part of a third party is not obvious, especially if it is destructive. Particular consideration is required by the subject of identification, which is determined based on the purpose of identification, and can have a rather diverse and peculiar manifestation. It all depends on who identifies the data subject, since the very concept of the data subject as a natural person with the totality of his attributes (direct or indirect) exists only in his understanding. For example, it can be the physical parameters of the body of the data subject, it can be information about the diseases of family members of the data subject, it can be information about the preferences of the data subject in food and clothing, it can be information about the daily routine of the data

subject, etc. In this case we can say, based on the definition of personal data, that identification of the data subject is the process of forming his information model of a given thematic focus like a dossier (file).

Let's start to study this question by considering a common scheme of the living environment of a modern natural person.

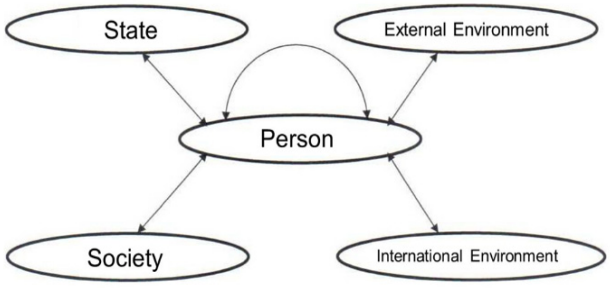

Figure 1. Common scheme of the living environment of a modern natural person

This scheme shows in general terms that in the course of his life, individuum (a natural person - a data subject) can enter into various types of relations with the state, society, the external environment, the international environment and another person (hereinafter referred to as objects of relationships).

In this case, the state is understood as the political form of organization of a society headed by the government and its bodies that manage the society, protect its economic and social structure, and society is a combination of historically established forms of joint activity of people in the broad sense of the word.

The external environment can be defined as a combination of the natural and technogenic environment of the natural person. The international environment is a set of forms of organization of the functioning of international communities, individual foreign states and their respective societies.

III. IDENTIFICATION OF THE DATA SUBJECT AND THE "DIGITAL TWIN"

Returning to the issue of identifying the data subject on the totality of information available about him, it should be noted that within the framework of the above scheme of relations between the individual and the objects of relationships, personal data is processed, during which direct or indirect identification of a specific data subject can be carried out. It should be borne in mind that the emerging relationship between the object of the relationship and the natural person can be one-sided, i.e. only one of the participants in this process is an active participant.

For example, the state in the person of the relevant structures can investigate the activities of the natural person in terms of tax payments, and the personality can explore the activities of another person in terms of the legitimacy of their sale of some property. In this sense, the identification of the data subject with destructive intentions in relation to the person is one-sided.

Direct identification of the data subject can be carried out on the basis of statutory signs (aggregated personal data) of such a unique identification. These signs may include:

- global identifiers such as the national identification number or any other identifier of general application (for example, electronic digital signature, SNILS in Russia or a personal bank code for individuals “henkilötunnus” in Finland);
- set of passport data (general passport, diplomatic passport, citizen's international passport, sailor's passport, etc.);
- set of data of various types of certificates (certificate of an employee of the Prosecutor's Office, certificate of a member of Parliament, certificate of a judge, certificate of a soldier, driver's license, etc.);
- combination of these registration documents (residence permit, temporary residence permit for stateless persons, invitation to enter the country, certificate of electronic signature, etc.).

With direct identification of the data subject, the determining set of personal data is stable and weakly dependent on temporal and spatial factors. On the contrary, indirect identification of the data subject may deal with a variable set of personal data related to temporal and spatial factors.

Indirect identification of the data subject is associative in the framework of these relations and is associated with various information that determines its physical, economic, financial, social, political, religious, medical parameters and various preferences, inclinations and interests. This identification can be carried out both independently and in conjunction with direct identification. In the general case, it does not require the presence of legally significant features of the data subject as a natural person, and in general with this type of identification, the data subject as an natural person may not be interested in who carries out this identification (relationship object), since in this case the primary subject of identification is the above signs.

This type of identification is supported by various types of analytical models and is the basis for constructing a digital profile ("Digital Twin") of a specific natural person. All further relations of an object of relationship with a natural personal can be built with its digital profile, since it does not need anything else to realize the objective function of the object of relationship in the framework of relations with the natural personality.

Indirect identification underlies the concept of Big Data, which is currently being actively developed in the field of personal data and is associated with the creation of a “Digital Twin” (“Digital Profile”) of a citizen. This raises the question of the need for anonymization or depersonalization of information about a natural person. This simplifies solution of issues with the processing of personal data, as laws of various countries on the protection of personal data do not consider anonymized information as personal data.

However, studies in this area show that supposedly anonymous information about a natural person allows to identify this concrete person with an accuracy of 99.8% if you will use Artificial Intelligence methods. So, the presence of 15 specific anonymous indicators of various

thematic focus allows researchers to almost uniquely identify a specific natural person. So, this anonymized information will be easy transferred into the category of personal data [6]. If we consider the Internet as a set of data that can be interpreted as Big Data, then the results of this study are interesting for personal data protection and security [7]. The study was given the task of identifying a natural person and his relatives by finding out their names, addresses, phone numbers, etc., using only one photograph of a student posted on the Internet. Within 15 minutes, the participants in the experiment, using search engines, pages of social networks, blogs and forums, found all the necessary information that allowed to identify the concrete natural person and his close relatives.

The purpose of direct identification is to correlate the data subject as a natural person with the legally relevant information (personal data) of a document in paper or electronic form, which allows you to implement the indicated relations of the object of relations with this person with a significant share of legitimacy. Moreover, the nature of these relations (constructive or destructive) is not important.

The purpose of indirect identification is more complex and can be both constructive and destructive. Indirect identification may be directed to:

- increasing the efficiency of the activity of the object of relationships within the framework of existing relations with the natural person;
- improving the quality of existing relationships and (or) the formation of new relationships for the natural person on the part of the relationship object and, as a result, increasing the efficiency of the relationship object itself;
- replacing the direct identification of the data subject with its indirect identification to conduct targeted impacts on the data subject (possibly a negative) having a physical, moral, material, financial or other nature, with the potential to change the relationship of this person with the objects of relationship;
- conducting targeted attacks to the digital twin of the natural person to violate his confidentiality, integrity and accessibility, which may result in the required change in the relationship of this natural person with the objects of relationship.

The basis of indirect identification are models and methods of data analysis that allow you to correlate associative data with the data subject. These models and methods determine the potential for effective identification of the data subject. However, its implementation depends on the purpose of identification, the associative data used and the methods of their processing, the capabilities of the relationship object, i.e. the identification process is probabilistic. In this regard, returning to the question of assessing the harm that may be caused to personal data subjects in case of violation of the law, it is necessary to apply a risk-based approach in identifying the data subject. In this sense, it is necessary to talk about the risk of harm, i.e. about the risk event and damage from its occurrence. In this case, the risk event is understood as the identification of the data subject in some context of its digital profile, and the damage is understood as some potential negative consequences of this identification (the risk of some destructive impacts on the data subject). For example, leakage of personal data as part of a full name, the national identification number or any other identifier of general application (direct identification) can lead to risks of negative impacts on the data subject in terms of his pension savings. Obviously, such risks can be somewhat applied to a specific the associative combination of personal data.

From the aforesaid, it follows that for information systems for processing personal data, along with models of threats and an intruder of personal data security, (these models are the source documents for organizing a personal data protection system) it is necessary to develop a model of consequences for the data subject in case of implementation of the actual threats identified in the threat model.

IV. CONCLUSION

At the present time digital transformation is being carried out in all areas of life of the modern state, society and the individual, including economic, social, political and other aspects. Digital models (“Digital Twins”) of physical objects, physical processes, decision-making processes, as well as natural person (the participants of these objects and processes) are created. A transition is being made from management at the level of physical objects, processes and natural persons to management at the level of their "Digital Twins".

The issue of the quality of these "Digital Twins" (adequacy, reliability, completeness, relevance) is at the forefront of security of personal data. If earlier various types of threats (negative impacts) were realized in relation to physical objects, processes and personalities, then in the context of Global Digitalization, negative influences are realized in relation to their "Digital Twins". After these influences, distorted (destroyed) digital models can has very strong impact to the physical environment of the modern state, society and the individual, while violating various types of security (included personal data security).

The issue of the quality of "Digital Twins" (even in the absence of negative impacts on them) is associated with factors such as:

- the amount of knowledge about the simulated objects, processes and their subjects;
- the presence of hidden structural and behavioral characteristics of the simulated objects, processes and subjects;
- own mistakes of digital modeling;
- the availability of various types of resources for digital modeling.

To reduce the influence of these factors, technologies of Big Data, Artificial (augmented) Intelligence, software and hardware robotics are used.

The development of ICT/Telecommunication technologies and software applications in various areas of the life of the modern state, society and the natural person generate a significant variety of many negative impacts on their digital models ("Digital Twins"). Moreover, at every current moment this set does not have a full description, which is reflected in the programming practice in the form of the thesis “the penultimate error is always

corrected". The potential for the implementation of these negative impacts depends on the specific conditions of use of "Digital Twin".

The described approach to the identification of the data subject leads to the idea that, in essence, the concept of the data subject in modern conditions begins to act as a digital complement (extension) of a natural person and his (her) environment, for which the digital profile ("Digital Twin") is synergistic. The characteristics of a digital profile can be synthetic concepts such as an index of intellectual development and may not be known to the natural person (the person himself knows nothing about himself). Therefore, the issues of the adequacy of the digital profile of a particular person ("Digital Twin") are of paramount importance, given the transition from physical reality to digital Augmented Reality (AR).

REFERENCES

1. Global Dat Protection Regulation. EU GDPR Portal. https://www.eugdpr.org.

2. Dokuchae V.A., Maklachkova V.V., Statev V.Yu. Identification of the subject is a key point in the processing of personal data. In the book: Technologies of the Information Society. *Proceedings of the XIV International Industrial Scientific and Technical Conference*. 2020. P. 273-274.

3. Hungarian overnment suspends EU data protection rights. Website "EURACTIV.com". [Electronic resource]. URL: https://www.euractiv.com/section/digital/news/hungarian-government-suspends-eu-data-protection-rights (accessed: 05/10/2020).

4. Dokuchae V.A., Maklachkova V.V., Statev V.Yu. Classification of personal data security threats in information systems. *T-Comm*. 2020. Vol. 14. No.1, pp. 56-60. (in Russian)

5. Doku aev V.A., Makarova D., Maklachkova V.V., Volkova L. Analysis of Data Risk Management Methods for Personal Data Information Systems. *2020 Systems of Signals Generating and Processing in the Field of on Board Communications.* 19-20 March 2020, 10.1109/IEEECONF48371.2020.9078538

6. Kommersant from 07.24.2019. Imaginary anonymity [Electronic resource]. - URL: https://www.kommersant.ru/doc/4040585 (accessed: 05/09/2020).

7. Stud ts in the debate on the protection of personal data managed to collect information about a person on the Internet on one photo. Roskomnadzor website, 12/03/2019 [Electronic resource]. - URL: https://https://pd.rkn.gov.ru/press-service/subject1/news4782/ (accessed: 05/09/2020).