DATA MODEL IN THE CONTEXT OF THE GENERAL DATA PROTECTION REGULATION

Tzanko Tzolov

Member of the Commission for Personal Data Protection
e-mail: tzolov@cpdp.bg
Bulgaria

Abstract: In 2013 The European Commission announced its Data Driven Economy initiative and started a number of procedures aimed at creating a digital single market. This concept represents a paradigm shift in the ways of describing each level of abstraction – economy, market, organization, process and it is a strategic direction in the development of the digital market. A formalization by model definition is needed and the DAMA Framework could help for this. The purpose of this article is to present an approach for abstract description of the data modelling. In this reason, metadata for different point of view are discussed. A personal data model under the Data Governance Model is designed and it is shown in conclusion.

Key words: Data governance organizations, business process, data flow, metadata, personal data model, GDPR

1. INTRODUCTION

In 2013 The European Commission announced its Data Driven Economy initiative and started a number of procedures aimed at creating a digital single market. Statisticians estimated that in 2015 the data-driven economy rose to 272 billion euros (5.6% annual growth) and could employ 7.4 million people by 2020. The "data-driven" concept represents a paradigm shift in the ways of describing each level of abstraction – economy, market, organization, process – and is a strategic direction in the development of the digital market. Data processing underpins this concept and is expected to be defined by using non-contradictory models.

With the help of the DAMA Framework [1] we can describe each organization using descriptors such as processes, technologies and roles. The model focuses on building capabilities in the 10 key areas that will allow the organization to achieve its business objectives.

Two types of processes underpin this model. The first type is business processes describing the achievement of objectives and the operation of the organization, and the second type is data processes transforming the abstractions into a series of information-related actions. The constraints change the input processing conditions or require the design of new processes.

This model can be applied to a variety of social relationships based on information (data) processing, information protection, and personal data protection. Moreover, the need for regulations on (non-)personal data has been increasing recently [2].

The transformation from business processes to data flows (processes) can be achieved using the definition of "data processing" in Article 4 of the General Personal Data Protection Regulation [3].

By developing the concept, data-driven organizations and the problem area defined as metadata management, we can present the organization through a model of its data. In such a model, the access to the data is limited by the status of the secondary data (metadata) that constitute data on the (real) data themselves. Users access the metadata that describe their status or actions related to them during processing.

For the purposes of the model, we are going to use an early classification of metadata [4] as metadata related to: (1) data discovery; (2) data availability; (3) data use; (4) administration and control [5].

The article will attempt to lay the foundations of a model where all constraints and parts of processes are described as metadata and supporting registers of certain events. The definitions of data processing set out in GDPR are going to be regarded as constraints.

2. ASSUMPTIONS

For the purposes of the personal data model, we are going to make the following assumptions:

- We are going to look at data governance organizations;
- The business intelligence organizations have thoroughly defined their data flows.
- Different types of personal data are divided into separate databases or are at least designated as such;

3. METADATA FOR DATA DISCOVERY. (A WAY OF RECOGNIZING PERSONAL DATA)

According to the definition set out in Article 4(1) of GDPR, personal data means any information that can directly or indirectly identify a person. This definition is not exhaustive; it is descriptive, and this empowers the controller to determine which personal data to process.

In order to determine the data processing regimes, we must be able to classify them by types that allow us to apply common processing policies. One such classification depending on the data processing regime is: (1) personal data; (2) special or sensitive data, and (3) public (personal data which have been made public by virtue of a law or another legal act).

Personal data themselves include categories (name, identification number, etc.), but also include location data, physical and physiological information, and psychological, economic and cultural data. Profiling, which refers to a person's conduct and behaviour, also falls within the scope of GDPR. For example, the fact that a person likes a particular letter or post on Facebook would now constitute personal data.

Here it is important to emphasize that a new collection of data whose elements have been defined as non-personal during collection can be treated as personal data during processing.

Special or sensitive personal data are referred to in Article 9(1), and a different processing regime has been created for them.

In order for the model to be complete, in addition to the types of data already mentioned we must also list data defined as (4) data with limited processing; (5) anonymized data; (6) archived data; (7) deleted data and (8) destroyed data. Types 5, 7 and 8 contain only metadata regarding the action, the reason for it, and the time of its execution.

It is important that organizations be able to identify the personal data they possess, to treat them separately from other data stored in their systems (financial, product information, etc.), and to classify them into a higher sensitivity level.

The recognition of personal data should not depend on the data carrier – electronic, paper or any other type of medium, i.e. depending on the type of processing used, we must have the necessary technology for recognition.

There must be a "Discovery" register or at least a data type flag.

4.METADATA FOR DATA AVAILABILITY

The conditions for personal data availability are linked to the compliance with the processing principles, which would make the processing itself and hence the availability of data to the controller admissible.

According to Article 5 of GDPR, personal data must be processed lawfully, fairly and in a transparent manner, collected for specified purposes limited to what is necessary, explicit, used for a particular period of time, and in a manner which ensures their integrity and confidentiality.

When personal data are being collected, some official information should be added to the records regarding: (1) the lawfulness of processing; (2) the purposes of processing; (3) safeguards that the data collected is limited to what is necessary; (4) the date of the last update; (5) the period for which they will be processed; (6) what protection measures will be taken during their processing.

The lawfulness of processing can be easily described by using the conditions set out in Article 6 and Article 9. This is a comprehensive list (within the meaning of the Regulation), and at least one of these conditions should be reflected in the system. The provision for lawfulness based on consent is of some interest. When information systems or online web-based applications are used, the requirement for a freely given consent is linked to: (1) the data subject's awareness; (2) consent should be given manually and not be deemed to be given by default; (3) the receipt of the service does not depend on the provision or non-provision of personal data. Child's consent is a special area governed by Article 8 and should be given by an adult who has custody of the child. From a technological point of view, this means that the controller should build capabilities to recognize the age and relationships of individuals. Given the possibility of repeated giving and withdrawal of consent, and in accordance with the principle of accountability, the personal data controller should create and maintain consents register with minimum information about: the subject, the purpose, and the status of the consent – given or withdrawn. The controller has to confirm the withdrawal of consent.

The principle of transparency requires that all information, both public and personal, should be concise, transparent, comprehensible, easily accessible, clear and unambiguous; also, when necessary, visualizations can be used. When addressed to the public, this information may be provided in electronic form, for example through a website. The collector

should also make it possible for data subjects to submit requests electronically, especially when personal data is processed electronically. The collector should respond to the data subject's requests without undue delay and no later than one month, as well as should state the reasons for refusing a request.

The principles of fair and transparent processing require the data subject to be informed and aware of the existence of such a processing activity and its purposes. The collector should provide the data subject with all necessary additional information in order to ensure the fair and transparent processing and inform him/her of the existence of profiling and the consequences of such profiling. When personal data is collected by the data subject, he/she should also be informed of whether he/she is required to provide personal data and of the alternatives in case he/she refuses to do so. This information can be provided in combination with standard icons so that the intended processing is meaningfully presented in a clear, visible, comprehensible and clearly legible way. If the icons are presented in electronic form, they should be machine-readable.

The processing purpose is determined either by the controller or can be legally regulated as an obligation. From a systemic point of view, the purposes should be formalized to a certain set of optional alternatives, leaving also room for the provision of additional information. The question of the admissibility of using personal data for purposes other than those originally intended is more interesting. In this case, a field for "further processing" purposes should be added and filled in accordance with the provisions of Article 6(3) and (4).

The safeguards for minimum amount of personal data necessary can be described as legally defined (if such a provision exists) or as a brief description of the processing algorithm with a focus on the personal data used.

The requirement for the data to be up-to-date is quite burdensome. Generally, it is the controller's responsibility, but the data owner is the one who changes data and renders them out of date. From this point of view, the controller has two options: (1) to look for public registers in order to update the information; (2) to enable the data owner to update his/her data by exercising the right granted to him/her under Article 16 of GDPR. In both cases, the record should contain the date on which the update was made and the manner in which the update was made. Obviously, in this case a history of the updates and the dates on which they were made must be kept. In order to reduce the volume of information stored, a compromise could be sought, such as keeping a record for a certain period of time or a certain number of updates if the law does not require the history to be stored in its full form.

The period for storing the data is determined either by a legal act or by the controller. GDPR introduces the "processing time" requirement – it is not possible to process data for an unlimited period of time. Alternatively, the time period can be stored as an absolute value (date) until processing is complete or as a period indicating how long the data will be processed after being entered.

The security measures for data processing will be laid down in a separate document drawn up on the basis of applicable data protection policies. These can also be descriptions (operational instructions) of the products and systems themselves, specifying the section related to personal data protection.

The last principle of "accountability" has a specific nature and is an attribute of the data controller or demonstrates compliance with all other principles and will therefore be presented separately.

5. METADATA FOR DATA USE

Following the data governance organizational model, each data category or register must have an owner and assigned access rights as well as, in some cases, access restrictions. This is part of the technical and organizational measures that the organization must take in order to ensure the protection of data.

In order to build an effective protection system it is necessary to establish and maintain roles within the organization that have different data processing rights. Information security is based on three types of data access:

discretionary access control (DAC) – a method based on identification and recognition; its distinctive feature is the possibility to transfer permissions;

mandatory access control (MAC) – a method based on levels of information criticality;

nondiscretionary access control (NAC) – a method involving centralized control and role-based or task-based control.

The DAC method is widely used in commercial applications. The administrator assigns permissions to users and grants them access to objects. DAC allows more flexibility in controlling who has access to what. It is the ideal method for implementing policies and system security management.

6. METADATA FOR DATA ADMINISTRATION AND CONTROL

Metadata for data administration and control ensure the accountability of actions constituting obligations of the controller or rights of data subjects. These actions relate not so much to the personal data themselves as to the compliance with the provisions of the Regulation. They can be regarded as administrative information on the data processed and pertain to the compliance with the accountability principle.

Articles 13 and 14 of the Regulation stipulate what kind of information should be provided to the data subject under the different data collection regimes:

We can generally divide data collection into two large categories (1) personal data collected by the data subject; (2) personal data collected from other sources.

When the personal data are collected by the data subject himself/herself, the personal data controller has to provide information that includes the minimum information listed in Article 13 of the Regulation. All information that has to be provided can be structured in a file and provided to the data subject during data collection – both in paper and electronic form. If the data are collected online, it must be possible for the document in question to be made available to the user upon request and prior to data collection completion.

It is logged in the system in the form of administrative information that such information is provided after the user has personally confirmed that he/she has read the information provided. The possibility of not providing this information if it has already been provided means that a register of the information provided must be maintained under Article 13 containing the following minimum information: date of provision; the number of the document provided.

When the information is collected from other sources, the personal data controller is obliged to notify the data subject by providing him/her with the minimum information set out in Article 14. The notification period is 1 month, at the time of the first communication to that data subject, or when the personal data is first disclosed to another recipient, whichever occurs first. In addition to the data under Article 13, the register of information provided should contain information on the reasons for provision or non-provision.

The two registers can be merged into one containing: the manner in which the personal data were acquired, the date of provision, the number of the document containing the information under Articles 13 and 14, the manner of provision/non-provision.

When the data subject cannot be provided with information on the origin of the personal data due to the fact that different sources were used, aggregate information is provided.

Under the Regulation the data subject is given significantly more rights to control the processing of data. This is related to the elaboration of administrative procedures and the provision of access to the information under Articles 15-22 of GDPR.

The data subject's right of access to the data under Article 15 of GDPR entails new requirements for the controller. Data access has two aspects – confirmation of the purposes for data processing, data categories, recipients or categories of recipients to whom data is disclosed, data retention period, the data source if the data is not collected by the data subject. The controller should also inform the data subject of his/her rights to rectify, delete and block the data, the right to file a complaint with the supervisory authority and whether the data is used as a basis for a machine-made decision. This information can be structured by the metadata already mentioned above and no special register has to be maintained.

If possible, the controller should be able to provide remote access to a "secure system" that will provide the data subject with a direct access to his/her personal data. This right should not adversely affect the rights or freedoms of other individuals, including trade secrets or intellectual property, and the software copyright in particular. However, these considerations should not constitute a basis for a refusal to provide all the information to the data subject. When the controller processes a large amount of information on the data subject, he/she should be able to ask the data subject to indicate the information or processing activities which are the subject of his/her request prior to data provision.

The second aspect is to provide a copy of the personal data that are currently being maintained by the controller. Technologically speaking, this does not constitute a problem, but in this case a register of the number of provided copies must be kept, insofar as it is acceptable that a reasonable fee can be charged after the first provision. A very important point is the fact that all personal data of third parties in copies of documents have to be deleted.

The right of rectification has been introduced in the form of a procedure whereby a data subject may request a rectification by providing the correct personal data himself/herself. The system should maintain a link to the document with the corrected data.

The right to be forgotten (RTBF) constitutes a technological challenge. Data controllers must ensure that the information related to the data subject were "deleted" not only from their systems but also from third party systems that have copied or are connected to the original information. In order to comply with this condition, a register should be maintained containing information not only on who has the right to receive the information, but also on the data actually provided to third parties.

The erasure of data is a debatable issue, as these data are also expected to be needed in the future in order to satisfy the public interest or in legal proceedings. Any retention would violate the rights of the data subject and it would be difficult to prove that a hypothetical need that may arise in the future is of an overriding interest. One solution is to restrict data processing, but additional safeguards must be put in place in order to ensure that the data are not used by the controller (even in this case this constitutes a violation if the existing provision is not amended). The data owner should be informed of this accordingly.

The problem with the need to maintain data for certain systems or rubrics related to other users, for example participation in forums or other data exchanges, is similar. In this case, deleting an account would change the mood and purpose of the whole discussion. In this case, the solution is to anonymize the data, but it would still be possible to recognize somebody based on the opinions and views expressed.

It is easy to see how organizations could become entangled trying to observe the letter of the law and to comply with RTBF.

It is also worth remembering that RTBF and other retention requirements are not absolute. Companies must decide whether the right of freedom of information can be affected, whether there are legal or public interest factors, or other concerns (Article 17(3)).

The right to restriction of processing may be used by the data owner in the cases referred to in Article 18. The controller may start to process again such data only after obtaining the consent of the data owner, and if the controller starts to process the data due to other reasons, he/she must inform the data owner. It is interesting that the restriction of processing may be used by the data subject where the controller no longer needs the personal data for the purposes of the processing, but they are required by the data subject for the establishment, exercise or defense of legal claims.

The methods of restricting the processing of personal data could include temporary transfer of the selected personal data into another processing system, suspension of users' access to them, or temporary removal of published data from a website. Data processing in automated personal data registers should in principle be restricted by technical means so that personal data will not be processed further and cannot be changed. The fact that the processing of personal data is restricted should be clearly specified in the system.

The data discovery and accountability conditions can be met by maintaining a register of restrictions and access to data whose processing has been restricted containing at least the following minimum information: the date on which the restriction was imposed, a document imposing the restriction, date of access, a document permitting processing.

The controller communicates any rectification or erasure of personal data or restriction of processing to each recipient to whom the personal data have been disclosed, unless this proves impossible or involves disproportionate effort. The controller should inform the data subject about those recipients if the data subject requests it.

In cases of consent-based or automated processing, the data subject has the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and has the right to transmit those data to another controller. Controllers should be encouraged to ensure the interoperability of the data format that allow data portability. This right should be exercised when the data processing of the data subject's personal data is either based on the data subject's consent or the necessity to perform a contract. The right should not be exercised when the processing is based on a legal basis other than consent or the necessity to perform a contract and does not create an obligation for controllers to adopt or maintain processing systems which are technically compatible. Where, in a certain set of personal data, more than one data subject is concerned, the right to receive the personal data should be without prejudice to the rights and freedoms of other data subjects. Furthermore, this right should be exercised without prejudice to the data subject's right to erase personal data and the limitations of that right, and in particular should not include the erasure of a data subject's personal data that he or she has provided under a contract, to the extent and for the time limits for which the personal data are necessary for the performance of this contract. Where this is

technically feasible, the data subject should have the right to transmit his or her personal data directly from one controller to another.

When the personal data are processed for direct marketing purposes, the data subject has the right to object free of charge and at any time to such processing, including profiling to the extent that it is related to direct marketing, irrespective of whether it relates to initial or further processing.

The data subject should have the right not to be the subject of a decision which may include a measure to assess personal aspects related to him solely on the basis of automatic processing and which produces legal effects for him or which affects him equally significantly for example, automatic rejection of online credit applications or electronic recruitment practices without human intervention.

Such processing includes 'profiling' that consists of any form of automated processing of personal data evaluating the personal aspects relating to a natural person, in particular to analyze or predict aspects concerning the data subject's performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements, where it produces legal effects concerning him or her or similarly significantly affects him or her.

Decision-making based on such processing, including profiling, should be allowed where expressly authorized for fraud and tax-evasion monitoring and prevention purposes conducted in accordance with the regulations, standards and recommendations of Union institutions or national oversight bodies and to ensure the security and reliability of a service provided by the controller, or necessary for the entering or performance of a contract between the data subject and a controller, or when the data subject has given his or her explicit consent.

In any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision.

The objections to processing should be kept in a separate register containing information on the nature of the objection, the date of the objection and the grounds for objection.

7. CONCLUSION

The presentation of the organization through a "data governance" model gives us a structured approach to introducing regulatory or ethical data processing restrictions. The restrictions are related both to the processing operations and to data availability and quantitative and qualitative data parameters. The use of different regulations requires data to be separated into individual databases, or at least into databases designated as personal and non-personal.

It is obvious that the data governance model is underpinned by the input and output data of business processes. The employees in the organization or its clients and partners receive roles with different rights or obligations to access data using available technologies for which the organization has built capabilities.

The personal data processed are regulated separately and should be recognizable as such, which is one of the great challenges. Recognition is not a one-time action, but an assessment process over the whole course of processing. The moment of data provision is particularly critical, as it should be taken into account that the user might have other data which combined with the data provided may make the identification of individuals possible.

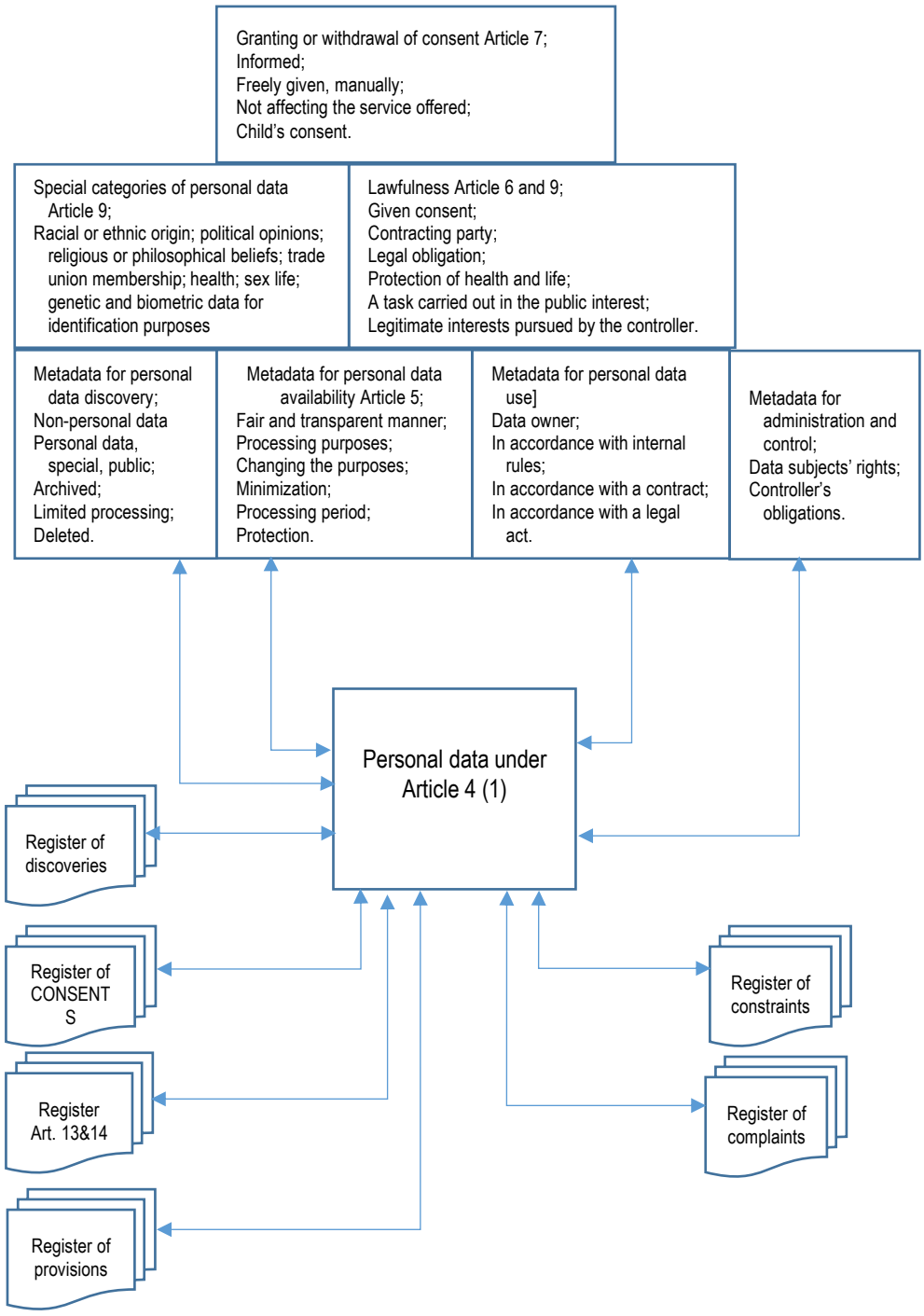

Fig.1 Personal data model under the Data Governance Model

The concept of using metadata can be considered from several perspectives. Users do not work with the data themselves but with their secondary characteristics, which speeds up processing and increases data protection. Using the data structuring model, we can design systems in line with the principle of privacy by design.

The structured data model (fig.1) can be easily aligned with changes in the legal framework by changing the metadata without affecting the data themselves.

REFERENCES

[1] Mark Mosley, DAMA-DMBOK2 Functional Framework, Data management Association, 2008;

[2] A framework for the free movement of non-personal data in the EU, Memo-17-3191_BG.pdf, Information document, European Commission, September 2017;

[3] General Data Protection Regulation, Official Journal of the European Union, 4 May, 2017

[4] Tedd, L., Large, A. Digital Libraries. Muenchen, 2005

[5] Intner, S., Lazinger, S., Weihs, J. Matadata and its impact on libraries. London, 2006,

Information about the author:

Tzanko Valkov Tzolov - member of the board Bulgarian Commission for Personal Data Protection. Master's degree: Automation Management Systems; National Security and Defense; Human Resources Management. Membership of professional bodies: Europol Joint Supervisory Body; Eurodac Supervision Coordination Group; Visa Information System SCGs Working Group 611.

Manuscript received on 15 October 2017