Business Process Modeling of a GDPR Compliant System for Research Project Management

[1]Raluca ARBA, [2]Adrian Vasile ARBA
[1]Babeş-Bolyai University of Cluj-Napoca, [2]3PillarGlobal, Romania
[1]raluca.arba@econ.ubbcluj.ro, [2]adrian.arba@3pillarglobal.com

***Abstract*–The digitization process and the wide use of Internet technologies have brought easy access to information and a significant improvement in the quality of life. At the same time it has also brought problems when dealing with privacy and personal data. European Union has issued the 679/2016 Regulation in order to set a legal framework for data protection and privacy. The law created a framework but the issue of translating this law into technical solutions remained the task of IT industry. The aim of this paper is to analyze the principles of data protection and translate them into a GDPR compliant model for research project management.**

***Keywords*: GDPR, privacy by design, privacy by default, research projects**

I. INTRODUCTION

The growing amount of information and the wide use of communication technologies have improved our lives but at the same time have led to a complex challenge given by the protection of personal data. Even if the concept of data protection is not something new, the introduction of the European Union's Regulation 679/2016 has made a significant change in the way data protection is taken into consideration and new models and architecture challenge the IT industry to focus more on privacy and security

The management of a research project involves teams of people working for different universities or research centers, a flow of personnel depending on the phases of a project, documents and personal data collected and processed for each member and shared with the other partners. This leads to the necessity of focusing on the issue of personal data protection, especially due to the flow of personnel that a project may encounter. Legal reasons are blended with specific requirements of the project and lead to the necessity of setting a consent for personal data collection and a system where traceability of personal data is a must. Most of the document management systems used are based on the workflow of documents not necessarily focusing on protection and traceability of data. The challenge is even more difficult when taking into account that the current data protection regulations come with a series of recommendations rather than solutions, leaving the technical details of implementation to the IT organizations.

The presented paper aims to analyze the concept of privacy by design and by default under the new European GDPR rules, identify the principles of data protection and propose a new model for research projects management that takes into consideration the above mentioned concepts. We aim to identify the particularities of a research project document and personal data collection, by implementing a flow register for collected documents, personal data and processed personal data and use this flow in a document management system.

II. LITERATURE REVIEW

.From the legal point of view, there have been many initiatives to handle privacy rights when processing personal data whether electronically or manual. The first convention was the one of the Council of Europe in 1981, followed by the Directive 95/46/EC of the European Parliament and the Council which granted EU citizens' rights for data protection and privacy. In April 2016 the directive was replaced by the new data protection regulations (Regulation 679/2016) also known as GDPR that came to effect in May 2018. The new law is applicable in all EU states and provided citizens more rights and introduced more constraints on organizations and controllers that process personal data in any purpose. [2],[3] One of the constraints introduced by the new regulation was the Data Protection by Design and by Default in the IT systems, constrained mentioned in the article 25 of GDPR. Penalties up to 4% of the annual turnover ore 20 Million Euro are punishment imposed in case of rule violation [4].

As stated in the article 25 of GDPR - "The controller shall implement appropriate technical and organizational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility."[4] Privacy by Design implies that privacy must be built into a system during the whole life cycle of it and companies must build any IT system that processes personal data with privacy and protection in mind. Privacy by Default means that once an IT system is built the best privacy settings are applied by default without the end user's intervention and data is kept for the amount of time necessary to provide the service or product.

The article offers general statements and rules and examples of techniques that may be implemented like pseudonymisation and encryption. It leaves to IT organization the freedom of implementing it properly [4].

DOI: 10.4316/JACSM.201902002

The difference comes from the fact that privacy by design was a good practice under the Directive 95/46/EC and becomes a legal requirement under GDPR.

The notion of privacy by design is not new. The concept was introduced in the mid 90' by Ann Covoukian to address the growing issues of the information and communication technologies [1].She introduced seven principles that became the foundation of Privacy by Design concept for many researches. ISO/IEC 29100 standard introduced 11 principles for privacy by design that focused on consent and choice, purpose, legitimacy and specification, collection limitation, data limitation, user retention and disclosure limitation, accuracy and quality, openness, transparency and notice, individual participation and access, accountability, information security and privacy compliance.

The literature [6], [7] review does not find consent on a certain number of principles to adopt, therefore we will focus on the GDPR and take into consideration the 6 principles stated by the current European law that handles data protection.

Under GDPR the following principles are mentioned:

- legality, equity, transparency
- limitations related to personal data scope
- integrity and confidentiality
- data minimization
- accuracy
- restrictions concerning storage

This leads to personal data mapping, simply translated into the need to understand the information flow, describe it and identify its key elements [8]. The challenge appears as each organization has its own system for data collection and processing, it deals with different processes and the need of a homogenous approach is becoming an obvious condition in order for the system to map the data correctly [9]. That may be translated into a specific model that leads to a system that unifies and gives research partners a way to keep their data inventory and maps, making it at the same time sharable with the other teams. The process of data mapping is presented in figure 1.

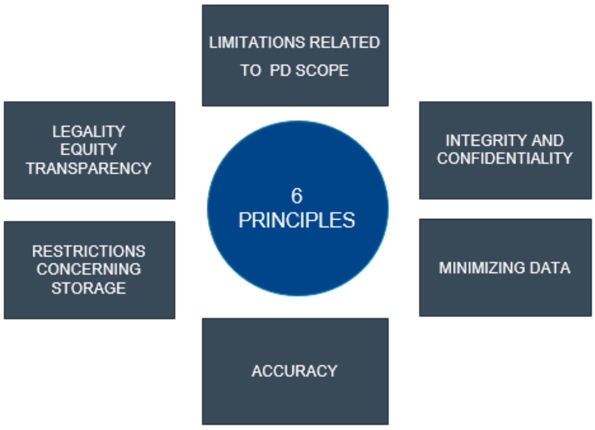
LIMITATIONS RELATED TO PD SCOPE
LEGALITY EQUITY TRANSPARENCY
INTEGRITY AND CONFIDENTIALITY
6 PRINCIPLES
RESTRICTIONS CONCERNING STORAGE
MINIMIZING DATA
ACCURACY

Fig.1 Principles of Personal Data Protection

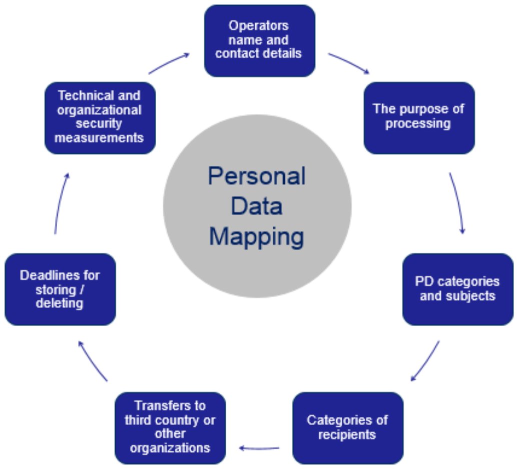
Operators name and contact details
The purpose of processing
PD categories and subjects
Categories of recipients
Transfers to third country or other organizations
Deadlines for storing / deleting
Technical and organizational security measurements
Personal Data Mapping

Fig. 2 Personal data mapping

Data maps offers organizations a visual or just a structured way to understand how data flows through its system. Organizations need to understand what kind of data are they collecting, how this information is used and who are they sharing the information with [10]. A wide range of tools from simple excel spreadsheet to sophisticated mapping programs are used to build data maps. Data compliance according to GDPR does not specifically mention the need to data mapping but it is considered best practice in order for organizations to fully have control and understanding of the personal data they are collecting and processing.

Taking all this consideration into account we have developed a methodology for data mapping that analyzes the data sources, the flow of documents and personal data through the stages of a research project administration, the internal and external entities that may have access to information and the main data operations that may occur and their documentation during the project development.

III. Methodology

We have started by analyzing several research projects from the first stage up to the moment when all the objectives were fulfilled and the project was declared finished. The document management of each project has been read in order to identify information that contains personal data and ways it was collected, used, processed and stored. We have divided this activity into three phases taking into consideration the type of information obtained.

The projects analyzed were the ones taking part in the grant competition PN II - Partnerships. This particular type of projects involve teams of universities working together on a research objective. One of the university is the coordinator and the other ones are considered partners. The partner submits the projects proposal with detailed information about each partner's contribution and is responsible for the general management of the project. Each partner is working on a particular task and collaborates with other teams in common tasks.

In the identification phase we have analyzed each project and have identified the documents and types of personal data that needed to be collected. We have divided this information into three main categories that need to be analyzed taking into consideration the main operations that can be performed: collect, update, modify and delete: collected documents (CD), collected personal data (CPD) and processed personal data (PPD).

The second phase was the analysis of this three categories and the perspectives from which this documents were collected. The following questionnaire was used to determine the flow of CD and CPD:

- **Who collects/processes documents/personal data?** When it comes to research projects documents and personal data are collected at the research department of the coordinating university and of the partners.
- **Is the data collected directly for the data subject?** Most of the collected data and documents are collected from the data subject or through forms at the process of project submission. Another moment when personal data is collected is the hiring process of researchers in a team.
- **Scope for collecting/processing documents/personal data.** The scope of collecting /processing personal data has to be correlated with the research activity and the data subjects involved. At the same time the scope for document and/or personal data collection has to be clearly stated and known by each member, regardless the team he is working for. The scope has to identify exactly the activities in which the collected documents and or personal data are used.
- **Is there a legal reason/ for collecting documents/personal data?** The legal reason has to be properly correlated with the scope and the activities in which personal data (PD) are used. According to EU's Regulation 679/2016
- **Internal procedures for documents and/or personal data collection**. Refers to the internal procedures of the research departments of each university involved. Universities have started to implement a data protection policy and are at different implementation stages of the GDPR policy. Even if each university has a set of internal procedures that describe the document management, those procedures were not focused on data protection and need to be adapted to the new data regulations.
- **Document/personal data storage format.** Usually documents are collected and stored in an electronic format as well as on paper. PD is collected through forms that team members fill in and is stored in an electronic format at the research department of each university.
- **Document / personal data is exchanged with other internal entities?** At the university level the documents and personal data are exchanged with different departments usually on an administration level. For example the IT department is involved when institutional emails are created for team members.
- **Document / personal data is exchanged with other external entities?** External entities the data may be exchanged with are the Ministry of Education and Research, The National Statistical Institute or the UEFISCDI Authority.
- **Period of time for storing documents/personal data.** Documents are stored for the duration of the project and archived afterwards. Storage of the documents and PD has to be properly managed with internal procedures that focus on data protection and clearly indicate where they are stored, who is responsible for them, what happens when documents and data are lost, etc.

Processed personal data (PPD) refers to the data that undergoes one of the following operations: use, update/ modification, data transfer, deletion.

- **The USE operation** refers to the use of personal data in daily activities of the project or of the research department of the university. We have taken into consideration which department is using the PD, in what format is the PD used and what is the reason for using that data.
- **The MODIFY/UPDATE operation** – we have taken into consideration what department may modify/update data, the reason for modifying/updating and where is the modification operated.
- **The DATA TRANSFER operation –** involves activities where PD are transferred from one university to the other or to external entities mentioned above. Data transfers should be realized according to internal procedures that focus on data protection.
- **The DELETE operation –** we have taken into consideration the following questions: the reason for data deletion, the moment of data deletion.

The analysis phase was the most important as it has identify reasons for document and personal data collection, legal laws that sustain the research activity and justify data collection and processing. It has identified the entities involved and the ways data is transferred and used by each one of them. The activity of data processing was divided into four categories with specific particularities.

The last phase involved the implementation of the analysis into flows of activities for the CD, CDP and PPD. A register of personal data was created that identified: persons involved in data collection and processing, activities that imply documents and personal data collection and processing, operations included in the definition of data processing and the ways they are used, internal and external entities that have access to data.

IV. RESEARCH PROJECT MANAGEMENT MODEL

The proposed model focuses on the document and data mapping process. The entities involved in the process are the partner universities, their research departments and team members.

The process starts when the personal data that need to be collected are identified. According to GDPR there are several categories of data: identification data and personal characteristics, details about family members, electronic identification, and financial element, other personal characteristics, physical characteristics, lifestyle information, hobbies and interests, affiliations, legal information, health data, information on studies and courses, information on professional history, information on political opinions, information on image processing, sound processing information, other types of personal data. On a research project level the following categories have been identified – identification data like name, surname, date and place of birth, email, signature, electronic identification, affiliations to university and professional organizations, financial details concerning card number, affiliated bank, lifestyle information like the functions a person have obtained, legal information, information on studies and courses, professional history, image and sound processing.

For each category of information there is an analysis whether this information has a legal reason for collecting and processing or does it need explicit consent. When consent is involved a common consent form that holds the minimal identification details is created and used throughout the project for each team and collected at the university level and shared with the coordinator and the UEFISCDI authority. Consent has to respect the GDPR rules be concise, explicitly state each category of data that needs to be collected, the reason for collecting, activities that will use the data, typed of operation that may be performed on data, the amount of time for data storage and use.

For the personal data that is only collected the flow will contain the following information: person/department responsible, reason, legal foundation for collection, and type of collection – electronical or manual, place and process of storage, time of data retention, deleting process.

Document collection is a step that usually happen in two moments – at the proposal of a project and at the moment of hiring a new team member. For the first moment documents are collected at the universities level and shared with the Coordinator and the UEFISCDI entity. At the hiring of a new research member the documents are collected at the university level, shared with at least the coordinator and the UEFISCDI authority.

The document flow is directly linked with the personal data flow as the second contains information obtained through the document collection process. At the university level the documents such as identity card, CV, passport, bank account statement are collected and stored during the time of the project and archived when the project is finished. From each one of those documents personal data is collected and entered into the research management system.

V. CONCLUSIONS

The process of mapping data into a system in order to be GDPR compliant is a complex task and raises challenges of adapting the existing systems to the new data protection rules. The means and technical solutions are left at the organizational level and IT systems.

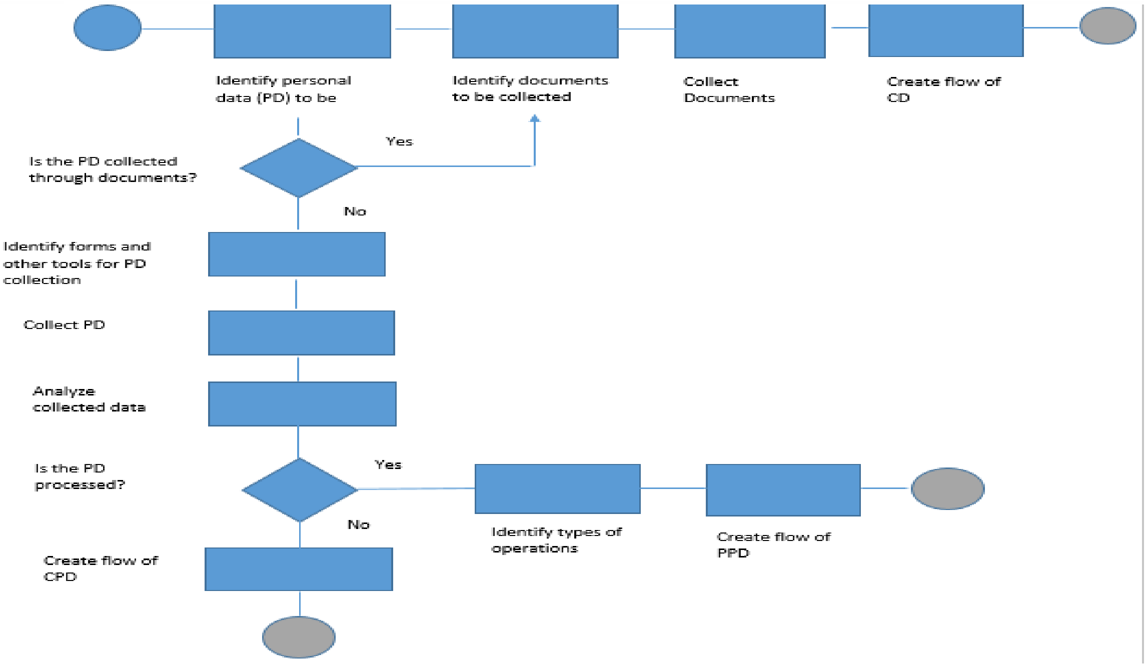
Identify personal data (PD) to be
Identify documents to be collected
Collect Documents
Create flow of CD
Is the PD collected through documents?
Yes
No
Identify forms and other tools for PD collection
Collect PD
Analyze collected data
Is the PD processed?
Yes
No
Identify types of operations
Create flow of PPD
Create flow of CPD

Fig. 3 Business Process Model for Document and Data Mapping

When it comes to research projects a considerable amount of information is collected and shared through different organizations and entities with the main scope of delivering a precise image of the research activity developed through the stages of a project and the actors involved. At the same time there is a challenge in keeping track and promptly responding to changes in the team and in the activities which usually leads to new personal data collection. Through a uniform system that maps data and documents correctly the task of data traceability should be manageable and effective.

References

[1]. Cavoukian, A. And Stoianov, A., 2007. Biometric Encryption. Biometric Technology Today, Vol. 15, No. 3, P. 11.

[2]. European Commission, 2011. Privacy And Data Protection Impact Assessment Framework For Rfid Applications, No. January, P. 1–24

[3]. European Commission, 2018, Data protection rules as a trust-enabler in the EU and beyond – taking stock. (COM/2019/374), July 2019

[4]. European Union GDPR 679/2016, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) available at: https://eur-lex.europa.eu/eli/reg/2016/679/oj

[5]. It Governance, Conducting a Data Flow Mapping Exercise Under the GDPR, GREEN PAPER, available at:https://www.itgovernance.co.uk/green-papers/data-flow-mapping-under-the-gdpr

[6]. Monreale, A., Rinzivillo, S., Pratesi, F., Giannotti, F., And Pedreschi, D., 2014 ,Privacy-By-Design In Big Data Analytics And Social Mining. Epj Data Science, Vol. 3, No. 1, P. 1–26.

[7]. Notario, N., Crespo, A., Martin, Y.S., Del Alamo, J.M., Metayer, D. Le, Antignac, T.,Kung, A., Kroener, I., And Wright, D., 2015. Pripare: Integrating Privacy Best Practices Into A Privacy Engineering Methodology. Proceedings - 2015 Ieee Security And Privacy Workshops, Spw 2015, P. 151–158.

[8]. Oetzel, M.C. And Spiekermann, S., 2014. A Systematic Methodology For Privacy Impact Assessments: A Design Science Approach. European Journal Of Information Systems, Vol. 23, No. 2, P. 126–150.

[9]. Schaar, P., 2010. Privacy By Design. Identity In The Information Society - Special Issue, Vol.3, No. 2, P. 267–274.

[10]. Vemou K.,Karyda M., An evaluation framework for privacy impact assessment methods,12th Mediterranean Conference on Information Systems (MCIS2018)At: Corfu, Greece, 2018

Raluca ARBA – is a lecturer at the Babeș-Bolyai University of Cluj-Napoca. Her research focuses on Business Process Modeling, Business Simulation, Web Semantics in the field of Ecommerce Systems

Adrian Vasile ARBA - is Compliance Manager and Data Protection Officcer at 3Pillar Global. His interests involve data protection and security, information security systems, quality assurance systems.