https://doi.org/10.31449/inf.v44i4.3253 Informatica **44** (2020) 497–505 **497**

Data Protection Impact Assessment Case Study for a Research Project Using Artificial Intelligence on Patient Data

Gizem Gültekin Várkonyi
International and Regional Studies Institute, Faculty of Law, University of Szeged
6720 Szeged, Tisza Lajos krt. 54, Hungary
E-mail: gizemgv@juris.u-szeged.hu

Anton Gradišek
Jožef Stefan Institute, Jamova 39, Ljubljana, Slovenia
E-mail: anton.gradisek@ijs.si

Keywords: data protection, DPIA, GDPR, Artificial Intelligence, medical data

Received: July 28, 2020

Advances in artificial intelligence, smart sensors, data mining, and other fields of ICT have resulted in a plethora of research projects aimed at harnessing these technologies, for example to generate new knowledge about diseases, to develop systems for better management of chronic diseases, and to assist the elderly with independent living. While the algorithms themselves can be developed using anonymized or synthetic data, conducting a pilot study is often one of the key components of a research project, and such studies unavoidably involve actual users with their personal data. Although one of the derogations stipulated in Article 89 of the GDPR is related to the data processed for scientific purposes, the GDPR still is applicable to that processing in a broader interpretation. The computer scientists and engineers working in research projects may not always be fully familiar with all the details of the GDPR, a close collaboration with a lawyer specialized in the European data protection legislation is highly beneficial for the success of a project. In this paper, we consider a hypothetical research project developed by an engineer dealing with sensitive personal data and a lawyer conducting Data Protection Impact Assessment to ensure legality and quality of the research project.

Povzetek: Prispevek obravnava varstvo osebnih podatkov pri uporabi metod umetne inteligence za analizo pacientov.

1 Introduction

In general, there are two ways to look at artificial intelligence (AI) dealing with personal data. On one hand, it offers great benefits for the users, if used correctly. For example, AI-enabled health care technologies could predict the treatment of diseases 75% better, and could reduce the clinical errors 2/3 at the clinics using AI compared to the clinics that do not [1]. On the other hand, the improper handling of personal data can quickly lead to abuse, sharing sensitive information, or other problems (unwanted disclosure, complex legal procedures, high amount of fines, etc.), therefore it has to be handled with the utmost care. In this paper, we will focus on the medical applications, such as the analysis of sensor data to help patients with chronic diseases manage their condition and improve the quality of life, or to help the elderly with independent living by providing safety features and improved communication channels.

Developing a product for the target population, for example people with diabetes, chronic heart failure, obesity, dementia, skin cancer, etc., typically starts with a research project, either in a company or within a consortium of research institutions and hospitals. One of the key components of such a project is collecting substantial amounts of data in a pilot study, with participants that resemble the target audience for the final product. When planning the pilot study, researchers enter a slippery terrain of dealing with personal data, as the participants are providing their own data for the purpose of the study. For the purpose of this paper, we will have a closer look at the medical data encapsulating three forms; general medical data provided by the medical doctor responsible for the participant, lifestyle data collected by either wearable or stationary sensors, and self-reported data that is obtained via questionnaires that the participants fill. At this stage, we only look at the data from the point of view of the hypothetical research project, and do not take into account the implications that are brought by potential commercial exploitation of the findings.

Right to data protection is one of the fundamental rights recognized in most of the European legislation, mainly in the Charter of Fundamental Rights and the General Data Protection Regulation (GDPR). The GDPR entered into force on the 25th of May 2018 replacing the Directive 95/46/EC based on two main aims: ensuring uniform data protection rights and rules EU-wide and towards data controllers, and keeping up with the technological developments challenging efficient

protection of personal data [2]. The effect of technology pointed out the need for more proactive ways to safeguard right to data protection and the GDPR mirrored this need by introducing a risk-based approach entrusted in the Article 35 of the GDPR introducing the Data Protection Impact Assessment (DPIA). As such, DPIA ensures data controllers comply with the GDPR requirements especially at an early stage of a new project. Those requirements could be specific to the right to data protection introduced in the GDPR such as the Article 25-Data Protection by Design, or to general principles that have already existed in European data protection legislation such as the principle of accountability. In fact, DPIAs are one of those ways for materializing and ensuring the accountability principle which has always been a legal compliance element and is now being utmostly challenged by the risks deriving from the new technologies [3].

The year 2018 was quite a productive year for the European Commission (EC) in terms of regulation of AI in the EU. Firstly, the EC published EU's AI Strategy [4] and then the EU's Coordinated Plan on AI [5] which both focused on the importance of system design which should be human-centric and trust-gaining. Both documents point out the data protection and privacy concerns as a problem, and suggest that legal compliance together with ethical system design is at the utmost importance to gain trust of AI users which then could boost the AI developments in the EU. The DPIA requirement embedded in the GDPR is such a tool that could be used as a proof before the users to gain their trust towards the AI system. For this reason, we think that the DPIA is an essential for any AI project planned to be targeted in the EU should be considered, if the project stakeholders aim at fulfilling both legal compliance and gaining individuals' trust. Individuals, who then might be data subjects in case they contribute to the AI system development in the training phase with their data, will enjoy the possibility of exercising their rights explicitly presented them by the DPIA output. Especially, they could receive descriptions specific to a systematic automated decision making processes since DPIA also aims to identify the logic involved in the algorithm as well as the significance of the consequences of the algorithmic evaluations [6]. However, yet there is no standard set for conducting a DPIA by the law-maker, as well as there is a lack of experience in practice since the GDPR is quite a young legislation. Specific to the AI driven research project collecting and processing personal medical data, there is no example existed in the literature illustrating a DPIA implementation, even though there are DPIA applications specific to AI implementations such as one for assessing the risks deriving from AI use in decision-making [7], or the work offering a roadmap for assessing the social and ethical impact of AI [8]. Furthermore, lack of specific examples to the DPIA on a certain technology, such as smart cities, may cause wrong identification of the risks which then may hinder data controller's full legal compliance [9]. In this paper, we aim to fill this gap with an example DPIA implementation on a hypothetical research project aiming to develop an AI system. Following content of the paper will be focusing on illustrating the legal foundations of the DPIA as in the GDPR in Section 2. Next, the hypothetical research case will be introduced which will then be followed by the DPIA practice in Section 3. According to the analysis conducted in the Section 3, the paper identifies three assessment titles specific to the AI projects: data specific assessment, data subject specific assessment, and project specific assessment. Section 4 presents a conclusion and a set of recommendations deriving from the outputs of the analysis conducted.

2 Data protection impact assessment in the GDPR

Article 35 of the GDPR does not provide an explicit description for the DPIA, however, Article 29 WP's guideline on the DPIA provides the following definition: "A DPIA is a process designed to describe the processing, assess its necessity and proportionality and help manage the risks to the rights and freedoms of natural persons resulting from the processing of personal data."

According to this definition, and in a narrower sense, the DPIA is a process consisting of several other sub-processes to describe the risks and assess the legality of the system in terms of data protection. These risks could be related to system security, system design, implementation, administration and development on a further run. The aim of the DPIA is to take appropriate safeguards to minimize the risks, if impossible to eliminate all. DPIA is not a simple one-time reporting activity, it is an ongoing process that should be continuously carried out during the lifetime of a project, therefore DPIA should always be monitored and updated [10].

It is the data controllers' responsibility to convey a DPIA, but in fact, the GDPR does not assign them an obligation to carry out a DPIA for every data processing activity. DPIA should be carried out when the data processing activity is likely to constitute a "high risk" to the rights and freedoms of natural persons (e.g. users of an AI service who both benefit from the service and contribute to it with their data), as the Article 35 (1) refers. The existence of automated decision-making tools applicable on personal data, and processing sensitive data such as medical data are some of the criteria conceptualizing the term high risk according to the Article 35 (3) of the GDPR. In addition, there are several guidelines published so far by the National Supervisory Authorities aiming to create a list of data processing activities likely to result in a high risk. Currently, all the National Supervisory Authorities of the 27 Member States have formed such a list. The European Data Protection Board assessed and delivered its opinions on each list to ensure consistent implementation of the rules in the EU. These lists could be the first sources for the data controllers to decide about the necessity of the DPIA for a certain project [11].

After determining the necessity to conduct a DPIA, the next step should be assessing the severity and likelihood of the risks which would come forward based on the data controller's own assessment. Although there is

no standard specified for how to convey a DPIA, failure to conduct a right assessment raises a risk for the data controllers; they may face several sanctions, especially financial penalties. Apart from that, conducting a right DPIA would be beneficial for the data controllers not only from the legal and the financial point of view. Wright [12] lists these benefits and refers that a DPIA could help data controllers to avoid implementing irrelevant solutions from the beginning of the project which may refer to assessing the technical feasibility of the system in parallel with the legal compliance. Therefore, the DPIA could help data controllers to save time and money. It also prevents the companies from losing their reputation (or from the scandals, as occurred with Cambridge Analytica, Equifax, Facebook, etc.). Finally, a DPIA document can be a trustworthy source which could be used as evidence before the public, and the related authorities to prove the data controller's respect to privacy.

When planning a research project, the DPIA shall not be conducted neither after launching nor during the implementation in order to ensure proactive measures. Specific to our project in the present article, the DPIA should be conducted based on two legal obligations as provided by the GDPR. Firstly, the Article 35 (3) point (a) of the GDPR clearly indicates that those data controllers that are using automated tools to evaluate personal aspects of natural persons, including profiling, are required to conduct a DPIA. Secondly, as the (b) point of the said article indicates, processing special categories of data also requires data controllers to conduct a DPIA. Medical data that will be evaluated in the project, as indicated before, is classified under the special data categories. Furthermore, the project focuses on developing an AI-based system that includes an automated decision making system (the AI software itself together with the algorithms to be developed) with profiling tools (surveys and hardware equipment). Based on these statements, it is clear that a DPIA must be conveyed by the data controller of the hypothetical project to see the risks and safeguard these risks in line with necessary tools. These tools might be either organizational or technical tools that could help mitigating the risks. The safeguards will be presented in the analysis part of the case study below.

3 Case study

In this paper, we present a step-by-step DPIA practice for our research project aiming to develop an AI-based healthcare software. While such approach is not a new in the literature and there are pieces of publications assessing the data protection risks of the real projects similar to ours [13], these are few in numbers and the DPIA really lacks in practice since data controllers usually do not prefer to publish their DPIA reports, as it is not required by law. We developed the idea of presenting a hypothetical research project that could exhibit a part of a realistic work and could contribute to the DPIA practices in the literature. Moreover, there are few examples specifically evaluating the data protection impact of an AI based software project. Following, we present the details of this AI software project and exhibit the simple DPIA elements that we created from several resources available on how to conduct a DPIA such as the Information Commissioner's Office guidelines[3], Deutsche Telekom's practices [14], French Data Protection Authority (CNIL) guidelines [15], and Article 29 Working Party guidelines [16].

3.1 Summary of the hypothetical research project

The goal of our research project is to discover new knowledge about a particular chronic disease and to develop a coaching system that will allow the patients for better management of their condition. In order to obtain sufficient amounts of data to build a personalized coaching tool, the researchers need to obtain various types of data from, say, 200 patients with the chronic condition, through a pilot study. In the pilot, the users are equipped with a smart wearable device that records the amount and intensity of daily activities, on aggregate. Such devices may come in the form of wristbands, smart watches, smartphones placed on various spots on the body, chest harness to monitor a simple electrocardiogram and breathing rate, or dedicated pendants. The interaction with the user likely takes place through a tablet or a smartphone. In addition, the application occasionally asks the patient questions related to their psychological state, as well as about some of their habits, such as smoking, consumption of alcohol, and about the dietary preferences. Medical doctors who recruited the patients for the pilots provide relevant data about the medical history of the patients, such as the timeline of their condition, the severity, and the medication or medical devices that the patient is using. The data is then processed using computer algorithms which find novel relations between various parameters, such as the effects of lifestyle on the expression of the condition, or how particular treatments help different patients best. Deeper information on the project will be gained during the DPIA, since it would be quite risky to first finalize the project details, and to conduct a DPIA afterwards [15].

Types of data subjected to the processing activity in the research project are qualified as special categories of data, also known as sensitive data, according to the GDPR. Sensitive data needs stricter protection, for example, the data subject's explicit consent should be obtained before launching the project. In order to obtain a valid explicit consent, the data controller must be able to present precise and specific information on the life-cycle of the data to be processed during the project. In addition, processing sensitive data may fall under the high-risk data processing category according to the provisions of the GDPR, therefore the data controller should conduct a DPIA prior to launching the project.

3.2 Requirements for a DPIA

The algorithm planned within the AI based healthcare software project is going to enable collecting data subjects' sensitive data based on profiling and processing that data. In addition, a large amount of data will be collected for feeding the algorithm conveying a risk for

data subjects, basically, the data subjects may cause loss of significant control of their own data. Based on these inputs, the project may reveal risks for rights and freedoms of the data subjects involved, if these risks are not mitigated. Therefore, it is a clear obligation for the data controller to conduct a DPIA and identify the risk categories with the planned mitigations when necessary. The following part shall present an assessment part of the actual DPIA since we skip the preparation phase of a regular process that includes planning, document collection, consultations with the stakeholders, etc. [10], which is not of particular interest for this paper.

3.3 DPIA for an AI-based healthcare research project

In this section, we conduct a DPIA on our research project. Several components are likely to be encountered while assessing any healthcare-related project, though each project has its own peculiarities. Therefore, this assessment is not universal, but can be viewed as an example for the engineers who are not deeply involved with the GDPR and who are looking for guidance to start with the preparation of the document.

We recommend that any DPIA should be conveyed under the supervision of a GDPR expert or a lawyer. As indicated before, the structure of the following section rely on several papers generated by the authorities guiding data controllers on how to conduct a DPIA. The questions and the answers referred to in this section stem from the author's own experiences, therefore the following DPIA is giving a lawyer's and an engineer's point of view. The structure of the below DPIA example is as follows: data specific assessment (DSA), data subject specific assessment (DSSA), and project specific assessment (PSA).

3.3.1 Data specific assessment

The DSA is chosen to be processed on the first hand because such an approach would shape the outcomes of the two other assessment groups. The DSA is the procedure where the data to be used in the AI project should be introduced very specifically in order to comply with the basic rules of the GDPR, mainly, the purpose limitation, transparency, accuracy, data minimization, and consent. It should be kept in mind that one of the requirements to be ensuring a valid consent is identifying the concrete data list together with the planned process of that data in the frame of a research project. Based on these statements, we propose the following questions placed in the Table 1. to be considered as part of the DSA.

The DSA questions raised here are related to the life-cycle of the data in the research project. Three types of data are planned to be collected during the research and all the types are clearly defined. The boundaries of the medical, activity, and self-reported data are reported in line with the data minimization and purpose limitation principles. Sources of the data are also clear and limited. It is crucial to note the responsible person for collection and processing of the data and the retention period is calculated. The data retention period should be followed without a prejudice to the Article 17 of the GDPR ensuring data subjects' right to erasure. This project does not aim at reusing data at the moment giving as a reason that there are several problems standing before data reusing rules and personal data protection legislation [17] refraining the project team from opening the research data for other purposes. However, there is a challenge identified with the models which will be reused, since it is well-known that with an intended attack, the data in training sets could be revealed [18], [19], [20]. This risk is mitigated with security measures and methods ensuring privacy specific within the AI models. In addition, AI models are not regulated under the GDPR except with the general rules such as Privacy by Design. More guidelines about protection of data in AI models could be delivered either from the European or from national data protection authorities. Finally, the 6th question in the table deriving from Article 13 and Article 15 of the GDPR raises concerns for the project team on how to ensure the full compliance with the GDPR if it is not possible to foresee the algorithm to be used from the beginning of the data processing. This problem is based on the technical construction of the AI and the legal uncertainty on the meaning of the logic-involved within the GDPR. If the term logic-involved means the planned algorithm to be developed, then it is not possible to give a concrete answer from the beginning of the project. If the rights vested in the Article 13 and Article 15 of the GDPR are the reactive rights rather than a proactive, then the project team can explain the logic of the algorithm, later. In any case, this risk is mitigated with a clear indication in the consent paper informing the data subjects about the concern.

3.3.2 Data Subject Specific Assessment

The DSSA should explain all the details of how the data controller ensures the rights of the data subjects and protects their informational self-determination right. The key point in this assessment is to gain trust of data subjects as required by law and ethics. The DSSA questions are mostly about how the data subjects rights will be ensured during the project. It is highly recommended to work with a Data Protection Officer in line with the Article 37-1 (b) of the GDPR, since data processing activities in this project require regular and systematic monitoring of data subjects on a large scale". The DPO whose duty is to consistently follow and ensure the communication between the data subjects and the project team, among the other tasks drawn in the Article 39 of the GDPR, could be chosen among the project team members, or to be contracted in line with the qualifications indicated in the GDPR. In this project, the DPO is the responsible person to guide the project team about the data subjects' requests and their fulfilment and is a lawyer specialized in data protection law. For this reason, we recommend the project team to work with a lawyer or a data protection expert from the beginning of the project development.

The importance of the 3rd question is vested in the clarity of the consent statement that shall be read and understood by each data subject. The project team could plan to involve the data subjects' opinion on the draft

1. What type of data, in what format, and in what scale will be processed?	Medical data (age, gender, medical history of other comorbidities, medications, clinical data related to the condition) Activity data (aggregated amount of physical activity per day and the physical activity, which is already defined in the course of the research and in the consent statement) Self-reported data (questionnaires prepared by the research team in collaboration with physicians) The study is limited to about 200 patients and the number will not exceed this.
2. What are the sources of the data? What measures are taken to ensure security of the sources of the data?	Medical data come from the treating physicians. Activity and self-reported data come from the patients through a smartphone application. All data transfers are secured with encryption algorithms and immediately anonymized.
3. Who will collect the data? Who will have an access to the data?	The medical data will be collected by their treating physicians. The activity and self-reported data will be collected through an application individually from each user. The access to the data will be structured hierarchically, with different partners having access to different types of the data, but only the treating physicians will know the identities of the patients, as this is unavoidable. All other partners will only access anonymized data.
4. Will the data be reused for another purpose in future?	The data will not be reused. What may be reused are the models that will be obtained by training the algorithms on the data. The models are protected with security and differential privacy measures against data revelation.
5. How long will the data be processed? Where and until it will be stored?	The data will be processed during the duration of the project, which is 3 years. It will be then stored for another 5 years for potential purposes related to the research within the scope of the project. It will be stored on a secure offline server physically located at one of the partner organizations. After that, all personal data will be deleted.
6. What technology will be used to process the data?	Before the data becomes available, it is difficult to answer this question. Typically, the algorithms used in such studies include decision trees, support vector machines, or different types of neural networks.
7. Who is responsible for the security of the data (in storage and during the collection)?	There is a dedicated engineer in the project team who is responsible for data security and storage.

Table 1: DSA questions for a DPIA and the corresponding answers regarding the project.

consent statement, and shape it in accordance with their feedback. Consultation with data subjects will also help the project team to know about their concerns, and mitigating their concerns would contribute the project to be more GDPR-friendly. The project team could also plan to make a half-day informative meeting with the data subjects on the project's essences and we present them the current DPIA. Data subjects are ensured with tools that could help them to withdraw their consent without an obstacle. Since the research is focusing on creation of the model, the users can ask for data deletion at any point while the development phase is ongoing. Once the models reach the final version, they will not contain any personal data. They are also given tools to download their data to be collected during the project and can exercise their right to data portability. It is planned that, since the development phase will take 12 months, the system will send automatic consent reminders every 3 months. Besides all these safeguards, data minimization is guaranteed with specific privacy setting interfaces embedded in the wearables or the sensors which will be used for lifestyle data collection. There will be a separate training designated for how to use the device and the privacy settings. The 4th question points out the well-known black-box debates that is weakening the intervention capability of the project team on the decision given by an algorithm, if the data subject wishes to exercise his or her right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision, as the Article 22 of the GDPR stresses. However, as this is a pilot project, the decisions are neither a final service nor a product, therefore they are only recommendations. Additionally, the project team guarantees the data subjects to express their own point of views (i.e. feedback) which can help the team to better develop or to modify the system mistakes.

1. What is the legal basis for processing data?	Article 6 (a) of the GDPR (data processing based on data subject's consent).
2. What is the purpose of the project?	The purpose of the project is to generate new knowledge about a chronic disease and to create a coaching platform that will be assisting the patients with management of their condition, thus greatly improving their quality of life.
3. What is the purpose of data processing? Are the purposes indicated in (2) fully in line with each other?	The purpose of the data processing is twofold: • to seek for new relations in the data which will lead to better understanding of the condition • to train personalized models that will make the experience for individual patients better
4. What is the expected benefit of the project for the data subjects, for the data controller, and for the society?	Overall benefit of the project will be for the individuals and for the society helping to develop a system that will be beneficial for them in future, for improving the quality of life, and reducing the burden on the health system for the society
5. How many stakeholders are involved with this project? Who is the data controller and how to identify the data controller?	If there are more than one stakeholders who can process the data, it is recommended to designate one of them as a contact point. The contact point's availability information should be easily accessible by the patient (address, phone, e-mail, preferable channel for communication, and the information of the DPO).
6. Are there any data processors? If yes, does the data controller have evidence of data processors GDPR compliance?	All personnel dealing with the data (data processors) are required to sign the GDPR compliance document, provided by the project leading institution.
7. How does the data controller ensure security of the data during collecting, processing, storing, and removing the data?	Collection: anonymization at the source, encrypted protocols for data transfer Processing: dealing with anonymized data only, no external access to the database Storing: offline storage at secure location after the project has ended Removing: authorized person deletes the data 5 years after the end of the project, if previously not requested by the patient Hardware safety measures, including wearables security, is ensured by the in-built safety measures of the devices provided by the manufacturer that proves GDPR compliance. The devices will communicate with the data-collection platform using only the patient's ID.

Table 3: PSA questions for a DPIA and the corresponding answers regarding the project

3.3.3 Project Specific Assessment

The PSA is the last part of our DPIA, presenting and explaining the legal basis for data processing, the project partners including the data controller, and the security measures that will be implemented to safeguard the data processed during the project. The security measures encompass a large and an important part of this assessment, therefore we present the security measures in a separate table.

The PSA table includes questions related to identification of the project purposes and legal basis as well as of the data controllers and processors. Data processing in this project is based on the data subjects consent and explicit consent where necessary. Data subjects are expected to participate in the project voluntarily, and they could make a decision about that participation based on the purposes of the project, expected outcomes and privacy statements to be provided for them. It is crucial for the project team to provide these information together with the explicit information on the identity of the data controller and other stakeholders, if there are any. In the PSA table, the second question raised some challenges while answering. The project aims at creating an AI-based healthcare software, however, since the data to be collected is Big Data, the project team is aware of the security risks and take necessary steps to ensure system security (see the Table. 4). Additionally, the project team ensures that, by design, the recommendation system will not work outside of the domain it was built for; therefore unpredictable outcomes are highly unlikely. In order to complete the administrative safeguards in this project, the project needs to present the necessary technical safeguards that are under the security measures.

Security risk	**Security measure**
Data partitioning (in relation to the rest of the information system)	Hierarchical access to the data, user roles
Logical access control	Hierarchical access to the data, user roles
Traceability (logging)	Use of dedicated file monitoring software
Integrity monitoring	Use of dedicated file monitoring software
Archiving	Periodic archiving
Paper document security	Paper documents kept in a locked filing cabinet or similar
General security controls regarding the system in which the processing is carried out	Dedicated preventive control measures
Operating security	Dedicated preventive control measures
Clamping down on malicious software	Up-to-date malicious software removing tools installed
Managing workstations	Dedicated personnel
Website security	Standard measures for website security
Backups	Periodic backups, automated
Maintenance	Dedicated personnel
Security of computer channels (networks)	Encrypted communication, when dealing with external sources (receiving the data during the pilots)
Monitoring	Data protection officer
Physical access control	Institute's physical access policy

Table 4: Security risks and measures (Extracted from [14], pp. 12-17).

The measures presented in the table are optional and may change depending on the project.

The final but an ongoing phase of the DPIA is the monitoring phase. Whenever there is a new element embedded in the project, and this element seems to change the balance of risk earlier assessed, the DPIA should be reviewed. This element could be involving a new data type in the algorithm or planning a commercial use of the algorithm. Bearing in mind the fact that ML and algorithms are referred to as entirely new technologies [3] and the growing amount of data together with a variety of hardware would raise the privacy risks [21], we suggest the project team to review the DPIA periodically.

4 Conclusion

Data Protection Impact Assessment is an integral part of any research project focusing on development of an AI algorithm with personal data. It should be conducted in the planning stage of the project and occasionally reviewed once the project is ongoing. This way, the project team, otherwise called data controllers, are able to identify the potential risks and find mitigation strategies for certain weak points. Last but not least, by conducting the DPIA, the project team fulfils the legal requirements, ensures higher trust of people involved, and avoids unforeseeable problems that might later occur.

It should be noted that there are automated general tools exist for the purpose of conducting DPIA [22], [23]. For instance, the open source DPIA tool freely offered by CNIL gives the possibility for the data controllers to compute the DPIA procedures with a step-by-step approach, and lets them make the risk calculation based on the weights identified for each risk labels, even though the risks are identified by the data controller manually. Users of the tool are guided with a set of questions categorized automatically and they could personalize the categories based on their needs. In the end, the tool gives the basics elements of the DPIA giving the data controllers opportunity to record also the mitigation records, but it would be highly beneficial for our project team to collaborate with a lawyer specialized in the data protection law, namely the GDPR, assisting them for using the tool. As demonstrated in this case study of a healthcare project

dealing with three types of personal data (medical, activity, and self-reported), DPIA is conducted through a series of steps, each of which addresses a different aspect of the data. We presented the DPIA in four tables, namely, Data Specific Assessment, Data Subject Specific Assessment, Project Specific Assessment, and Security risks and measures. Each table focuses on the particular aspect of the project and as close as it is compliant with the legal requirements. Step-by-step approach helped us to divide the data processing procedures and then evaluate each in detail. At the end of this assessment, it is safe to state that there is a low level of risk in this project, from the data protection point of view. This study aims to be an example for the community who is to plan an AI project and is looking for practical prior guidance to conduct a DPIA.

In this case study, we only focused on the use of personal data for the purpose of the research project. Clearly, successful research projects often continue with follow-up studies and eventually lead to commercial systems, in our case for example a smartphone-based coaching application for better management of a chronic disease. Commercial exploitation of research resulting from analysis of personal data opens a new series of questions. Can we commercially exploit the outcomes of this research project? Can a potential coaching application developed during the project be licensed to a commercial partner that will offer a subscription-based service? How to cover this in the agreement form that the participants sign before the beginning of the pilots? Such questions will be addressed in a future study.

Acknowledgement

AG acknowledges the funding from the ERA PerMed project BATMAN, contract number C3330-20-252001. On Slovenian side, the project is funded by the Ministry of Education, Science, and Sport (MIZŠ).

References

[1] "The AI effect: How artificial intelligence is making health care more human", [Online], study conducted by MIT Technology Review Insights and GE Healthcare, 2019. Accessed from: https://www.technologyreview.com/hub/ai-effect/ Last accessed: 20 April 2020.

[2] EDPS (2012). "Opinion of the European Data Protection Supervisor on the data protection reform package", (7 March 2012).

[3] ICO (2018). Accountability and governance: Data Protection Impact Assessments (DPIAs).

[4] European Commission (2018). Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, Artificial Intelligence for Europe. COM (2018) 237 final.

[5] European Commission (2018) Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, Coordinated Plan on Artificial Intelligence. COM (2018) 795 final.

[6] Kaminski, M. E. and Malgieri, G. (2019). Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations. International Data Privacy Law, 2020, forthcoming, U of Colorado Law Legal Studies Research Paper No. 19-28, Available at SSRN: https://ssrn.com/abstract=3456224 or http://dx.doi.org/10.2139/ssrn.3456224.

[7] Ivanova Y. (2020) The Data Protection Impact Assessment as a Tool to Enforce Non-discriminatory AI. In: Antunes L., Naldi M., Italiano G., Rannenberg K., Drogkaris P. (eds) Privacy Technologies and Policy. APF 2020. Lecture Notes in Computer Science, vol 12121. Springer, Cham. https://doi.org/10.1007/978-3-030-55196-4_1.

[8] ECP Platform for the Information Society Netherlands (2019). Artificial Intelligence Impact Assessment. Last accessed: 21 April 2020. https://ecp.nl/wp-content/uploads/2019/01/Artificial-Intelligence-Impact-Assessment-English.pdf.

[9] Vandercruysse, L., Buts, C., Dooms, M. (2020). A typology of Smart City services: The case of Data Protection Impact Assessment. Cities, 104, 102731. https://doi.org/https://doi.org/10.1016/j.cities.2020.102731.

[10] Wright, D. (2012). The state of the art in privacy impact assessment. Computer Law & Security Review, 28(1), 54–61. https://doi.org/https://doi.org/10.1016/j.clsr.2011.11.007.

[11] EDPS, Data Protection Impact Assessment: Accessed from: https://edpb.europa.eu/our-work-tools/our-documents/topic/data-protection-impact-assessment-dpia_en Last accessed: 20 April 2020.

[12] Wright, D. (2011). Should Privacy Impact Assessments Be Mandatory? Commun. ACM, 54(8), 121–131. https://doi.org/10.1145/1978542.1978568

[13] Horák, M., Stupka, V., & Husák, M. (2019). GDPR Compliance in Cybersecurity Software: A Case Study of DPIA in Information Sharing Platform. In Proceedings of the 14th International Conference on Availability, Reliability and Security. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3339252.3340516.

[14] Deutsche Telekom (2018). Guideline For the design of ai-supported business models, services and products at Deutsche Telekom in compliance with data privacy regulations.

[15] CNIL (2018). Privacy Impact Assessment Templates Accessed from: https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-2-en-templates.pdf Last accessed: 20 April 2020.

[16] Article 29 Working Party (2017). Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679.

[17] European Commission (2020). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A European strategy for data. COM (2020) 66 final.
[18] Aïvodji, U., Gambs, S., Ther, T. (2019). GAMIN: An Adversarial Approach to Black-Box Model Inversion. ArXiv, abs/1909.11835.
[19] Melis, L., Song, C., Cristofaro, E.D., Shmatikov, V. (2018). Exploiting Unintended Feature Leakage in Collaborative Learning. 2019 IEEE Symposium on Security and Privacy (SP), 691-706.
[20] Giuseppe A. et al. (2013). Hacking Smart Machines with Smarter Ones: How to Extract Meaningful Data from Machine Learning Classifiers. arXiv:1306.4447v1.
[21] Chandra, S., Ray, S., Goswami, R. (2017). Big Data Security: Survey on Frameworks and Algorithms, in 2017 IEEE 7th International Advance Computing Conference (IACC), Hyderabad, pp. 48-54. https://doi: 10.1109/IACC.2017.0025.
[22] CNIL Open Source PIA software tool. Accessible here: https://www.cnil.fr/en/open-source-pia-software-helps-carry-out-data-protection-impact-assesment. Last accessed: 20 April 2020.
[23] Alnemr R. et al. (2016) A Data Protection Impact Assessment Methodology for Cloud. In: Berendt B., Engel T., Ikonomou D., Le Métayer D., Schiffner S. (eds) Privacy Technologies and Policy. APF 2015. Lecture Notes in Computer Science, vol 9484. Springer, Cham.