

A Tool to Support Automation of Risk Assessment

Mauro Iacono[1], Michele Mastroianni[2(✉)], Francesco Palmieri[3], and Antonio Saporito[3]

[1] Università degli Studi della Campania "L. Vanvitelli", viale Lincoln 5, 81100 Caserta, Italy
mauro.iacono@unicampania.it
[2] Università degli Studi di Foggia, Via A. Gramsci 89/91, 71122 Foggia, Italy
michele.mastroianni@unifg.it
[3] Università degli Studi di Salerno, Via Giovanni Paolo II, 84084 Fisciano, SA, Italy
francesco.palmieri@unisa.it

Abstract. In the context of risk analysis and assessment, the constant evolution and growth of organizations has led to an increase in the amount of information to be analyzed: a case is provided by the data processing registry, mandatory according to the GDPR. Quality and cost of risk analysis and assessment may be improved by using tools to automate certain steps. In this paper we propose an approach and a support tool for software-aided risk assessment integrated with the state-of-the-art privacy impact assessment (PIA) tool, to define how certain steps in the compilation of a DPIA (Data Protection Impact Assessment) can be automated. The resulting tool chain has been successfully applied to a case dealing with job placement data in a university.

Keywords: Privacy · GDPR · Compliance · DPIA · PIA Tool · Record of Processing Activities · Risk Analysis · Risk Assessment · Information systems

1 Introduction

As organizations have continued to grow and become more complex, there has also been a significant increase in the amount of information they have to examine; consequently, it becomes essential to carry out risk analysis and assessment procedures on an ongoing and recurring basis. In a number of business settings, therefore, the fact that risk assessments and analyses are carried out manually rises as a major issue. The increasing complexity and volume of data has led to a search for new approaches aimed at enhancing the efficiency and accuracy of such assessments [9]. This paper proposes to address this issue by means of a tool capable of automating specific steps in the process of risk analysis and assessment, focusing on data management in the context of the Record of Processing Activities[1], with particular attention to the University Job Placement

[1] According to art. 30 of GDPR.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
O. Gervasi et al. (Eds.): ICCSA 2024 Workshops, LNCS 14822, pp. 126–143, 2024.
https://doi.org/10.1007/978-3-031-65318-6_9

section, thus inherently going to transform a process-based data structure into an asset-based one.

The chosen approach involved an in-depth study of a risk analysis and assessment tool widely used in business contexts: the Privacy Impact Assessment (PIA) tool, developed by the Privacy Authority of France. The main aim of this paper is to propose a toolchain in order to automatize some steps in the redaction of a DPIA (Data Protection Impact Assessment) carried out through the PIA tool. We analyzed the existing theoretical and practical basis for risk analysis and assessment and the role of risk analysis and assessment in cybersecurity, checking the different approaches to risk assessment and the importance of continuous management. A focused overview of GDPR, the Data Processing Registry and DPIA is provided in this paper, focusing particularly on the university environment and the Job Placement section to allow a correct connotation of the PIA tool with respect to the case study, with a detailed analysis of their functionality and role in conducting the Data Protection Impact Assessment, in order to show the reader the main issues in the design and development of our approach and of the related toolchain.

With respect to standard procedures, the implementation of the tool designed to automate the compilation of a DPIA provided promising results, reducing compilation time and improving accuracy through the reduction of human errors.

This paper is organized as follows: Sect. 2 introduces the main points about risk analysis and assessment, Sect. 3 resumes the relevant aspects of the GDPR, Sect. 4 presents the DPIA methodology, Sect. 5 discusses the PIA tool and its role in the toolchain, Sect. 6 describes the design process, conclusions close the paper.

2 Risk Analysis and Assessment

Risk analysis and assessment in the context of cybersecurity are key processes for identifying, understanding, and mitigating threats to the security of data and information systems. One of the best ways to address cybersecurity issues is to take a risk-based approach. When it comes to real-world implementation, information security risk management is a challenging process because risk factors can constantly change and evolve. Information security risk assessment (ISRA) is an important part of an information security management system (ISMS) and allows an organization to identify potential threats and decide what prevention or mitigation measures to adopt [11]. Risk management is an ongoing, recurring process that all organizations should do. Organizations that do not conduct risk assessment properly and regularly can suffer serious damage and consequences such as loss of reputation, legal problems, or financial impact.

Several ISRA approaches have been developed over the years, Some of approaches have been conducted using well-known risk assessment methodologies such as the Failure Mode and Effects Analysis (FMEA) [12], while other approaches are founded on methodologies particularly suited for the ICT field, such as NIST SP 800-30 [10] for further information, a comprehensive literature

review about privacy risk assessment methodologies may be found in [14]. However, the problem with many of these methodologies is that they focus primarily on general principles and guidelines, leaving users without adequate implementation details. The goal of risk assessment is to identify all possible risks to resources and accurately assess them to adequately mitigate them. Therefore, the vulnerabilities of each valuable resource (i.e., assets, services, business processes) and the threats that could exploit them are to be identified, and then the relationship between vulnerability and threat is defined as risk. Existing current approaches to risk assessment and information security are classified according to three criteria [8]:

- **Quantitative:** This type of approach involves the use of numerical data and statistics to assess risks. Typically, historical and probabilistic data are used to calculate the probability of a damaging event and its consequences in terms of financial loss, property damage, damage to reputation, etc. Quantitative approaches often involve the use of mathematical models, simulations, and statistical analysis to estimate risks very precisely. The inputs and outputs of quantitative risk assessment can be classified into two categories: monetary and non-monetary. As for monetary assessment, a monetary value is assigned to each asset, vulnerability, threat, and safeguard implementation. In contrast, non-monetary assessment produces a non-monetary number. Quantitative valuation is also known as "expected value analysis."
- **Qualitative:** This type of approach, unlike the quantitative approach, is based on subjective assessments expressed in qualitative terms. Rating scales or matrices are often used to score risks based on factors such as likelihood and impact. Experts in the field assess risks more descriptively, without precise numerical calculations. This approach is useful when quantitative data are lacking or when a much faster and more general assessment of risks is desired. The inputs and outputs of qualitative risk assessment can be classified into two categories: range and linguistic variables for inputs, and range and range (or rank) variables for outputs. In the range variable category, which applies to both input and output, the qualitative assessment evaluates risks in a qualitative risk matrix, which includes a measure of vulnerability and a measure of threat probability, each of which is expressed on a scale of three to five levels (e.g., low, medium, and high). In the category of linguistic variables, expert opinion is used to weight criteria and evaluate alternatives. The concept of linguistic variable is very useful in dealing with situations that are not well defined.
- **Hybrid (Semi-quantitative):** This approach combines both quantitative and qualitative approach methods. Quantitative data are used when available, but supplemented with qualitative assessments when data are scarce or uncertain. For example, you can use a numerical scale to express the probability of an event, but rely on expert assessments to estimate the impact [4]. This approach offers a trade-off between the accuracy of numerical data and the flexibility of qualitative assessments.

2.1 Information Security Risk Management

Information security concerns the protection of information of value to an organization from possible exposure, manipulation, or unwanted destruction. The ultimate goal of an information security program based on risk management is to maximize the organization results (products or services) while minimizing what are the unforeseen negative effects from potential risks. It is critical that risks are managed in such a way that instills confidence in all parties involved and ensures an adequate level of security for the information systems that support the day-to-day activities of organizations [1]. Through proper risk management, a balance between potential and acceptable risks can be achieved. Nowadays, there are different types of frameworks for managing information security risks. Each framework is created to meet specific needs and, as a result, has different objectives and steps. A general information security risk management framework includes four main processes, namely risk definition, risk assessment, risk response and risk monitoring [10]:

1. **Risk definition:** this is the initial stage in which potential information security hazards are to be identified and clearly defined. This includes identifying possible security threats, vulnerabilities in systems, or sensitive data that could be compromised. The goal is to understand what are the risky problems or situations that could jeopardize information security.
2. **Risk assessment:** once all potential hazards have been defined, the next step is to assess them. At this stage, an attempt is made to determine the probability and impact of each identified risk. The likelihood goes to measure how likely it is that a particular hazard may occur, while the impact goes to measure the consequences or damage that may result from the occurrence of the hazard. This assessment allows one to go in and determine which risks are the most critical and require the most attention. There are four ways to address a given risk:
   - **Accept**: The organization understands the risk and its consequences and consciously decides to accept it;
   - **Prevent**: The activity that exposes the organization to one or more risks is avoided altogether;
   - **Transfer**: All or part of the responsibilities and burdens associated with a particular activity and its associated risk are transferred to another party;
   - **Mitigate**: The risk and its consequences are controlled and limited in some way by reducing the risk to a level below that accepted by the organization.

   Usually, an organization risk management involves a combination of these ways: some risks are avoided, some are transferred, some are mitigated, and the others are accepted. As part of the risk assessment process and based on the organization policies, objectives and stakeholder interests, the level of risk acceptance (criteria) must be defined.
3. **Response to risks:** after assessing potential hazards, strategies must be identified and developed to address them. This may include implementing

security measures to mitigate or reduce risks. Risk management aims to control or mitigate possible threats to information security. Information security risks that exceed the level of risk acceptance must be addressed in one of the other ways of risk response, namely by avoiding, transferring, or mitigating the risk (mitigation).

4. **Risk monitoring**: the risk monitoring process is an ongoing process and includes constant oversight of identified hazards and implemented security measures. This ensures that preventive actions are effective in the long run and that new potential hazards are detected in a timely manner. In addition, risk monitoring allows changes to be made to risk management strategies as threats and conditions evolve.

One of the most important aspects of the risk assessment process is the perspective on which it is conducted. Three different perspectives can be considered [11]:

- **Asset-driven**: in this type of perspective, threats and vulnerabilities of each asset are identified and risk scenarios are formed. In a medium to large organization, the number of information assets is considerable. These assets represent valuable resources within the organization. Each information asset is associated with a variable number of risk scenarios, i.e., potential situations in which these assets could be compromised or damaged. The result of this multiplicity of assets and risk scenarios is the creation of large lists of risk scenarios that require careful analysis and constant maintenance. This is the main disadvantage of this perspective, which makes the risk assessment process a tedious and error-prone task. The main advantage, on the other hand, is that they are easier to understand, and given their widespread use, most of the tools available on the market are designed based on this perspective.
- **Service-driven (service-oriented)**: in this type of perspective, all important services in the organization are identified, and risks are identified and assessed according to their impact on them. Unlike asset-driven approaches, this perspective identifies threats and vulnerabilities for each service and not for individual assets. Because the number of services in an organization is significantly less than the number of assets, a service-driven approach is easier to manage.
- **Business-driven**: this type of perspective is based on the goals and processes that support an organization. Values are assigned to business processes linked to business objectives because these processes are more understandable and easily estimated than information assets. This perspective makes it possible to have a risk assessment that is much easier to explain to top management and thus acquire the necessary support.

In addition, risk measurement is the step in which the level of risk associated with a given scenario is quantified. This process is essential for determining whether existing security measures are adequate or whether additional mitigation actions are needed. Two types of risk measurement can be distinguished:

- **Unpropagated**: the risk of a single event or scenario is considered without regard to possible consequences or impacts on other events, processes or assets within the organization. In essence, risk is assessed in isolation, and this is useful in understanding the risk associated with a specific threat or vulnerability.
- **Propagated**: considers not only the initial risk, but also how this risk might propagate through the organization. It focuses on the possible chains of events and related impacts that could occur as a result of the initial risk. This type of measurement is much more complex and takes into account the interconnections between various business processes, systems, and activities. This type of measurement is more comprehensive and better reflects how threats can have knock-on effects within the organization. However, this type of approach requires more analysis and understanding of business dynamics, whereas unpropagated risk measurement provides a more focused view of specific risks.

The choice between these two types of measurements depends on the complexity of the organization and the objectives of the risk assessment. In many cases, it is advisable to consider both measurements to obtain a complete risk assessment.

3 GDPR and Record of Processing Activities

GDPR is a European Union regulation on the processing of personal data and privacy, aimed to strengthen the protection of personal data of European citizens and residents, both within and outside the borders of the European Union. This Regulation was drafted with the aim of protecting the fundamental rights and freedoms of individuals, particularly the right to protection of personal data. According to the European Commission, personal data are all the information inherent to an individual and therefore can cover any type of data such as names, addresses, bank data, health data, and more. However, the GDPR regulates only data related to individuals (including data in the professional or associational sphere), excluding data related to entities with legal personality such as public bodies, associations, etc. The types of data defined by the GDPR are divided into:

- **Personal data**: is information about an identified or identifiable natural person;
- **Genetic data**: are the data related to the genetic characteristics of an individual and provide unique information about the physiology or health of the natural person;
- **Biometric data**: are the data related to the physical, physiological or behavioral characteristics of an individual and which enable or confirm their unique identification (facial image, fingerprints, etc.);
- **Health-related data**: are data related to a person's physical or mental health.

One of the cornerstones of the GDPR is the Record of Processing Activities (RPA), which represents an electronic document with the aim to collect in a systematic way all operations conducted within an organization to process collected data. The importance of the RPA cannot be underestimated, as the GDPR explicitly requires organizations to have one. Every row of the RPA is structured in several fields:

- **Details about the organization**: Name and contact details regarding the person in charge of data processing and of the data protection officer.
- **Purpose of data processing**: Describes the purposes for which the data collected are processed. An example would be corporate personnel management;
- **Categories of personal data**: Defines all categories of data collected e.g. name, contact details, financial data, health data etc.
- **Data recipients**: represents a list of people to whom data may be disclosed or transferred such as service providers.
- **Retention periods**: Indicates the length of time the information will be retained before it is securely stored or deleted. This may vary depending on the purpose of the processing or the type of data collected
- **Security measures**: indicates what technical and organizational measures the organization has taken to protect data, this could include the use of encryption, backup procedures, etc.
- **Breach log**: documents data breaches, actions taken to resolve, and notifications to data protection supervisory authorities.

The Record of Processing Activities must be complete, accurate and constantly updated, and it needs to be easily accessible to data protection authority and people involved in the data processing process within the organization. The information contained in RPA have to be considered of paramount importance, and could form the basis for a risk assessment procedure [3].

3.1 The Entities Involved in the Processing

The GDPR also defines who is involved in processing on different bases:

- Based on the purposes for which the data are collected:
  - **Data Controller**: is the person or organization that collects personal data and decides why and how such data will be processed;
  - **Joint Controller**: is the person or organization that shares the same data processing purposes with another Co-Processor and together they decide how the data will be processed;
  - **Data Processor**: is the person or organization that processes personal data on behalf of the Controller and this is done on the basis of a contract or legal agreement between the Controller and the Processor;
  - **Recipient**: is the person or organization that receives personal data from the Controller and processes it according to the instructions received. In some cases, the Recipient may itself become a Controller for the data received if it processes it for its own purposes;

  - **Authorized Person**: is an individual who has received precise instructions from the Controller to carry out specific data processing activities;
  - **Data Subject**: is the natural person to whom the personal data collected and processed refer. A Data Subject provides his/her data to a Data Controller for purposes specified in the Privacy Policy. The Data Subject also has the right to have control over his/her personal data and to exercise his/her rights under the Data Protection Act;
- Based on the characteristics of the Owner/Director and the types and amount of data processed:
  - **The Data Protection Officer (DPO)** is a legal figure responsible for overseeing all aspects of GDPR implementation on behalf of the Data Controller or Processor. This figure is mandatory for public administrations and ensures compliance with data protection laws by helping to ensure that the rights of data owners are respected.
- Based on the territorial scope:
  - the Representative in the Union of the Data Controller/Responsible Person who is based in a third state is the legal or natural person who on behalf of the Data Controller/Responsible Person acts as a point of contact for data subjects and the Union Supervisory Authorities. This does not absolve the Owner or Manager from overall responsibility for data processing;
  - the Supervisory Authority is a public legal entity established by each member state to oversee implementation and compliance with the GDPR within its territory. It plays a key role in ensuring compliance with data protection laws;
  - the European Data Protection Board is a legal entity at the European level tasked with coordinating the work of the various Supervisory Authorities and supporting the European Commission in promoting and implementing the GDPR. It helps ensure consistency and uniformity in the application of data protection regulations throughout the EU.

The definition and distribution of roles and responsibilities is important in order to develop appropriate security measures for the compliance to GDPR. In addition, to prevent, avoid and minimize risks, the GDPR has introduced a methodological approach called risk-based i.e., a risk-oriented approach. This approach is strictly based on protecting user data and assessing the actual risk for each organization. Risk assessment thus becomes crucial in the context described by the Regulation. The GDPR, in fact, requires the Data Controller and the Data Processor to put in place the necessary technical and organizational measures to ensure a level of security appropriate to the risk, based on its assessment. Hence they must take into account: i) the state of the art of security measures/technologies; ii) the related costs of implementation; iii) the nature, scope, context of data; iv) the purposes and risks of processing.

3.2 Institutional Purposes of the University

In Italy, universities are considered legal entities with public legal capacity and pursuing purposes of general interest. They operate under the jurisdiction of

administrative law and have public authority; therefore, they are considered part of the public administration under Italian law. According to the Constitution, universities enjoy educational, scientific, organizational, financial and accounting autonomy, and this autonomy is regulated by their statutes, internal regulations and relevant legislative regulations. The main purposes of data processing that universities carry out are teaching and research. As for teaching, the main activities related to it include, for example, ensuring the right to study, guidance, curricular courses, tutoring, etc. As for research, on the other hand, universities are involved in national and international research projects and the publication of scientific articles. In recent years, however, the so-called "Third Mission" of the university has emerged as an additional institutional purpose. This mission is also concerned with the dissemination of knowledge and technologies outside the institution itself and includes areas such as:

- technology transfer (patents, third-party services, spin-offs);
- lifelong education (training on the working-age population, organization of conferences, conventions, etc.);
- social engagement (job placement, startup assistance, etc.). o social engagement (job placement, startup assistance, etc.).

4 DPIA: Data Protection Impact Assessment

A data protection impact assessment or DPIA (Data Protection Impact Assessment) is a particular procedure designed to identify, assess and manage the risks associated with a specific type of processing. A data protection impact assessment is required in situations where processing may pose a high risk to the rights and freedoms of individuals. This assessment is mandatory in at least three specific circumstances:

- An accurate and comprehensive analysis of the personal aspects of a person, including profiling;
- The processing of sensitive data on a large scale;
- The systematic and large-scale monitoring of public spaces.

Impact assessment is a process that must be conducted prior to the initiation of processing and should be conceived as a dynamic process, rather than seen as an isolated activity. As mentioned in the previous Section, the GDPR adopts a risk-based approach, which implies that the data controller must manage all potential risks associated with the personal data it processes. This requires the identification, analysis, estimation and mitigation of risks through the implementation of specific security measures [13].

4.1 Why Conduct a DPIA?

A Data Protection Impact Assessment (DPIA) can be conducted for any advanced or innovative procedure related to the processing of personal data.

In some cases, it may be mandatory to conduct a DPIA in accordance with applicable legal regulations. Indeed, personal information can be valuable to the organization that processes it, but it is important to recognize that the processing of such information carries significant liability because of the potential risks to the privacy of the individuals involved. Personal information, however, also has important meaning for those to whom it belongs. Indeed, this data can serve administrative, commercial, or even help build one's image. However, it is critical to recognize that security breaches in data protection can result in negative consequences, causing physical, material and moral harm. Given the often high stakes and evolving systems and threats, risk management allows you to determine what controls are needed to be in place. It also allows for methodical study of the processing of personal data, prioritizing risks and treating them proportionately to optimize costs and make decisions based on the most objective information possible [7].

4.2 How to Conduct a DPIA

In order to conduct a successful Privacy Impact Assessment (DPIA), the following steps must be followed [2] (Fig. 1):

1. **Activity planning:** Creation of a working group, definition of general and detailed objectives, planning of activities, including tasks, time and resources;
2. **Information Gathering:** Acquisition of all necessary information and data through interviews, document collection, and consultation with the offices and departments involved;
3. **Data analysis and description of processing:** Conduct an analysis of the data collected to identify and describe the purposes, categories of data subjects, data involved, and legal basis for processing;
4. **Data Flow**; Use of graphical descriptions (diagrams) of data flow for better assessment of critical processing profiles;
5. **Identification and risk assessment:** Identification of critical issues in terms of risks to the rights and freedoms of data subjects. Assessment of necessary measures to eliminate or reduce these risks in accordance with the principles of the *GDPR*;
6. **Review of assessments made**. Possible review of the analysis and assessments by those involved.
7. **Implementation of mitigation measures:** Implementation of the identified mitigation measures designed to manage and reduce the risks revealed during the analysis and assessment.

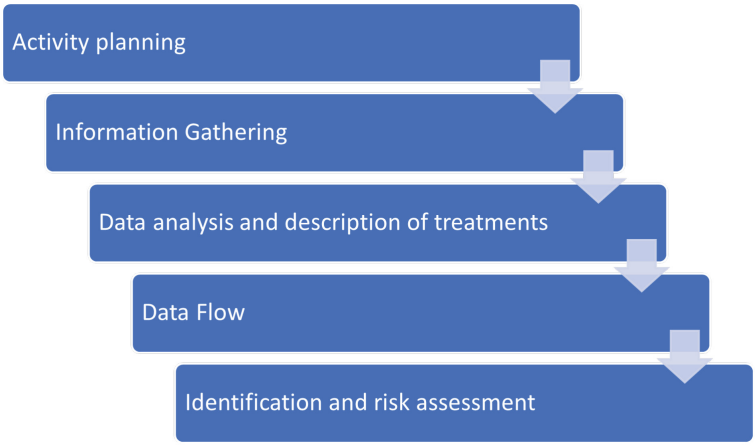


Fig. 1. Steps for conducting a DPIA

5 PIA Tool (Privacy Impact Assessment)

When conducting a DPIA, all steps performed must be documented in detail and accurately by the data controller. The final report will be sufficient evidence, if necessary, of compliance with applicable regulations, in line with the principle of accountability/accountability. The final DPIA document will be signed by the data controller. One of the most relevant tools used for redact a DPIA is the PIA tool, developed by French Data Processing Authority (CNIL) and available for download in CNIL site[2]. This tool allows you to have a preset structure in which to enter the different information for each stage of the DPIA. In fact, the tool is divided into sections, where for each section there are a series of questions to be filled in order to proceed with the various steps to conduct a DPIA [6]. At the end of the process of filling in the information, a document can be extracted that must contain at least the following parts:

1. Introduction: This section provides an overview of the personal data processing analyzed, introducing the key concepts and objectives of Privacy Impact Assessment (DPIA);
2. Body of the DPIA:
   (a) Scope Description: This section goes on to specify the scope and limits of personal data processing, providing a clear understanding of the activities involved.
   (b) List of legal controls: this section specifies and explains the legal controls that guide the processing of personal data, ensuring regulatory compliance.

[2] https://www.cnil.fr/en/open-source-pia-software-helps-carry-out-data-protection-impact-assessment.

(c) List of risk treatment controls: The various controls implemented to mitigate risks associated with the processing of personal data are identified and described, emphasizing the focus on security and privacy protection.
(d) Risk map: A visual representation is used to illustrate the risks associated with the processing of personal data.

3. Conclusion: This section provides a justification for the validity of the DPIA, going on to highlight the findings and decisions made during the assessment process.
4. Appendices:
   (a) Detailed description of the scope: this section further describes the details of the scope of personal data processing.
   (b) Detailed presentation of controls: this section sets out in much more detail what controls are implemented, with an explanation of their functions and purposes.
   (c) Detailed description of risks: this section provides an in-depth view of the identified risks, including potential consequences and strategies to mitigate them.
   (d) Action plan: this section goes into specific actions to be taken to manage the risks and improve the security of personal data processing over time.

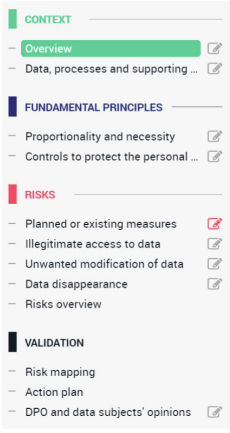


Fig. 2. Sections of the PIA tool

6 The Proposed Solution

The design of the proposed tool has been carried on by applying a process consisting of the phases which follow.

6.1 Collection of Information for DPIA Automation

The first step in this work was to analyze the information provided by the PIA tool. Starting with the graphical interface of the tool, it was seen that it is divided into several sections, each of which provides a series of questions. For each question, one should indicate an answer inherent to the type of analysis to be conducted. Once you have completed entering the information, the tool allows you to produce a final PDF document for conducting the DPIA [5]. In addition, the tool also allows the DPIA to be exported to a JSON file, which will be the focus of this work. In fact, the basic idea was to go and reuse this JSON file to make a tool to automatically compile the DPIA and then import this already compiled file to the PIA tool. In order to do this, the first thing done was to create an empty DPIA and export its JSON file, and then we went on to analyze this file. As shown in Fig. 2, the file will contain a set of information related to the DPIA (such as name, creation date, Author, etc.) and a set of parameters related to the DPIA.


```
"pia": {
    "status": 0,
    "is_example": 0,
    "is_archive": 0,
    "created_at": "2023-11-15T08:30:38.673Z",
    "name": "DPIA Saporito Antonio",
    "category": null,
    "author_name": "Saporito Antonio",
    "evaluator_name": "Mastroianni Michele",
    "validator_name": "Palmieri Francesco",
    "updated_at": "2024-01-16T08:45:56.204Z",
    "structure_id": "",
    "id": 19,
    "progress": 22.5
},
"answers": [···
],
"measures": [···
],
"evaluations": [···
],
"comments": []
```


Fig. 3. JSON file of an empty DPIA

In the context of analyzing this JSON file, we focused on the following fields:

- the answers field, which will contain all the information related to the questions proposed by the tool, with their answers;
- the measures field: which will contain all the information related to the existing measures regarding the DPIA to be conducted.

As shown in Fig. 2, the fields listed above, turn out to be empty due to the fact that an empty DPIA was exported, consequently the next step was to look

for a way to be able to identify the various questions that the PIA tool makes available and have them included in the JSON file. This was done by using hexadecimal codes to be inserted into the text boxes for the answer fields of the various questions, so that they would show up in the JSON file and consequently then be able to locate them.

Specifically, the idea was then to take advantage of a hexadecimal code structure to identify the questions and their fields to be filled in on the JSON file, as shown in Fig. 3. Since the tool is structured in sections (or Divisions), where each section contains a set of questions whose answers are to be entered, for the hexadecimal code structure it was thought to make a code divided into two parts and composed of:

- question section: the first two characters of the hexadecimal code, for example D1 as shown in Fig. 3, represents the first section. The D character was chosen to identify the section or Division in which the question is located;
- question number: the last two characters of the hexadecimal code, for example as shown in Fig. 3, R1 represents the first question contained in the section. The R character was chosen to identify the Response field to be filled in.

At this point, after the hexadecimal codes for all the fields requested by the tool were entered, the JSON file was exported and it contains all the fields requested by the tool with the corresponding hexadecimal code entered manually. For the measures field, however, hexadecimal code could not be entered because there are no associated forms with text boxes to fill in. Therefore, in order to automate this part as well, the JSON file containing at least one manually entered measure was parsed, so that the necessary parameters could be identified in the code to proceed with the automatic entry of measures. Once all the necessary fields were identified in order to be able to automate the process of compiling the DPIA, thought was given to how to provide the JSON file with the necessary information for compilation. Regarding the answers field and thus the entry of answers for each question, the basic idea was to make a CSV file structured that will be filled in with the relevant answers to be transferred to the JSON file (Fig. 4).

The Record of Processing Activities related to a particular data processing activity (in this case, Job Placement) and the Chat-GPT artificial intelligence tool was used in part to compile the answers for each question.

6.2 Python Code for DPIA Automation

At this point, once the CSV file was structured, thought was given to how to transfer the information, contained in it, to the JSON file of the PIA tool. To do this, a source code written in Python was developed through the Google Colaboratory development environment. As a first step, it was planned to make a code that takes as input the JSON file provided by the PIA tool and compiled with the entered hexadecimal codes. The following is a step-by-step description of the code:

id_question	question	text
D1R1	What is the processing under consideration?	text of the answer
D1R2	What are the responsibilities linked to the processing?	text of the answer
D1R3	Are there standards applicable to the processing?	text of the answer
D2R1	What are the data processed?	text of the answer
D2R2	How does the life cycle of data and processes work?	text of the answer
D2R3	What are the data supporting assets?	text of the answer

Fig. 4. CSV file containing the tool questions with the respective codes

- Definition of a function that checks if the file extension is of type JSON, if not, it prints an error message;
- JSON file upload request;
- Checking the file extension: If it does not have the correct extension, prompts the user to upload the file again;
- Extraction of JSON content;
- Loading the contents of the JSON file into a variable called date;
- Initialization of an empty list that will be used later to save responses to be updated;
- Printing an upload confirmation message.

At this point, a second code snippet was developed, very similar to the previous one, but it goes to take as input the CSV file containing the information to be passed to the JSON file. Once the CSV file has been uploaded, the information present in it will be extracted and made suitable for later use. After having loaded both the JSON file and the CSV file and having read the information present in the latter, the next step was to develop a code snippet that allows you to transfer the information extracted from the CSV file and insert it into the JSON file in the appropriate fields.

Below is a step-by-step description of the code:

- Check the presence of the 'answers' response field in the JSON file: if the field is present, the code proceeds to modify the responses in the JSON;
- Iterate through the responses present in the JSON file and the lists containing the information extracted from the CSV file;
- Print question and answer information;
- Modify the answers by adding specific information: concatenate the answers taken from the CSV file to the original text, in this case the question identifiers;
- Provides feedback in case the number of lines in the CSV is insufficient to match all the responses in the JSON;
- Print a message if the 'replies' field is not present in the JSON.

Once the process of transferring information from the CSV file to the fields in the answers field was finished, the next step was to automate the process of inserting existing measures for the measures field. To do this, an additional code snippet was created, below is a step-by-step description of the code:

- While Loop: this cycle repeats infinitely until it is explicitly interrupted by the user, it allows you to continue inserting measures;
- Retrieving the PIA ID, if it is not present, the null value is assigned;
- Check the presence of the 'measures' field. If present, the user is asked to provide the title and content of a new measure;
- Creation of a New Measurement Record;
- Adding the "Measure" to JSON file;
- Confirmation message of the addition of the new measure;
- Control of the Continuation of the While loop: The user is asked if he wishes to insert another measure, if the answer is negative, the program ends, otherwise we proceed with inserting another measure.

The last code snippet involves filling in two more sections of the DPIA by going for the IDs (identified in the JSON file) related to the subsequent questions. To do this, an additional CSV file was made that contains the question ids related to the sections to be filled in and the related question. For each value in the column called "id_reference_to," it looks to see if an object with the same value exists within the "answers" field in the JSON file. If the object exists, it displays information about the object (question id, question, list of answers) and asks the user if he or she wants to add or edit a list associated with that object. If the answer is "y," the user can add new items to the list, which is then updated in the JSON file. If the object does not exist, the snippet provides for the creation of a new object, including the insertion of items into the response list associated with that object. The new object is then added to the JSON file. Once the compilation of the DPIA has been completed, it is necessary to be able to save the new JSON file which will then be the file to be imported into the PIA tool in order to complete the compilation of the fields where human intervention is necessarily required. To be able to save the new JSON file, a specific code snippet was created.

The proposed toolchain has been tested using real data of the job placement administrative process of the University of Campania Luigi Vanvitelli already used in other works [3], and the detailed results, the map of codes and the related CSV are available on request from the authors.

7 Conclusion

In conclusion, in this research work, has been developed a tool to (partially) automate the process of Risk Analysis and Assessment in university processes, in this case the Processing of Personal Data for Job Placement purposes has been assessed. Automating those processes offers several benefits to organizations, including:

- **Efficiency and Speed:** Automation allows for faster and more efficient analysis and risk assessment than manual methods. Repetitive and detailed tasks can be handled faster, enabling timely decision making.

- **Reduction of Human Errors:** Human errors are inevitable, but they can be reduced through automation. Eliminating or minimizing human errors in the analysis and risk assessment stages contributes to greater accuracy and reliability of information.
- **Managing Large Quantities of Data:** Many organizations must manage huge amounts of risk-related data. Automation facilitates the analysis of large datasets, enabling the detection of patterns and trends that are difficult to detect manually.
- **Resource Savings:** Automation allows human resources to focus on more complex, value-added tasks, reducing the workload associated with more repetitive tasks.

As far as future developments are concerned, regarding the automation process carried out for the PIA tool, one could think of going for an additional CSV file regarding the existing measures, so as to create a kind of Dataset containing all possible mitigation measures and let the user dynamically choose which measures to add to the DPIA without having to manually enter text.

Acknowledgements. This research was partially funded by MUR PRIN PNRR 2022 grant number P20227W8ZC, and by project SERICS (PE00000014) under the NRRP MUR program funded by the EU - NGEU.

References

1. Brunner, M., Sauerwein, C., Felderer, M., Breu, R.: Risk management practices in information security: exploring the status quo in the dach region. Comput. Secur. **92**, 101776 (2020). https://doi.org/10.1016/j.cose.2020.101776
2. Campanile, L., Cantiello, P., Iacono, M., Marulli, F., Mastroianni, M.: Risk analysis of a GDPR-compliant deletion technique for consortium blockchains based on pseudonymization. In: Gervasi, O., et al. (eds.) ICCSA 2021. LNCS, vol. 12956, pp. 3–14. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87010-2_1
3. Di Martino, B., Mastroianni, M., Campaiola, M., Morelli, G., Sparaco, E.: Semantic techniques for validation of GDPR compliance of business processes. In: Barolli, L., Hussain, F.K., Ikeda, M. (eds.) CISIS 2019. AISC, vol. 993, pp. 847–855. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-22354-0_78
4. Ferra, F., Wagner, I., Boiten, E., Hadlington, L., Psychoula, I., Snape, R.: Challenges in assessing privacy impact: tales from the front lines. Secur. Privacy **3**(2), e101 (2020). https://doi.org/10.1002/spy2.101
5. French Data Protection Authority (CNIL): Privacy Impact Assessment (PIA) - Knowledge Bases (2018). https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-3-en-knowledgebases.pdf
6. French Data Protection Authority (CNIL): Privacy Impact Assessment (PIA) - Methodology (2018). https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-1-en-methodology.pdf
7. Henriksen-Bulmer, J., Faily, S., Jeary, S.: Dpia in context: applying dpia to assess privacy risks of cyber physical systems. Future Internet **12**(5) (2020). https://doi.org/10.3390/fi12050093, https://www.mdpi.com/1999-5903/12/5/93

8. Landoll, D.: The security risk assessment handbook: A complete guide for performing security risk assessments, 3rd Ed. CRC Press (2021)
9. Mastroianni, M., Palmieri, F., Ficco, M., Kozik, R., Choraś, M.: Privacy risk analysis and metrics in capturing and storing network traffic. In: 2023 24th International Conference on Control Systems and Computer Science (CSCS), pp. 580–585 (2023). https://doi.org/10.1109/CSCS59211.2023.00097
10. National Institute od Standards and Technology (NIST): Guide for Conducting Risk Assessments (2020). https://www.nist.gov/privacy-framework/nist-sp-800-30
11. Shameli-Sendi, A., Aghababaei-Barzegar, R., Cheriet, M.: Taxonomy of information security risk assessment (isra). Comput. Secur. **57**, 14–30 (2016). https://doi.org/10.1016/j.cose.2015.11.001
12. Spreafico, C., Russo, D., Rizzi, C.: A state-of-the-art review of fmea/fmeca including patents. Comput. Sci. Rev. **25**, 19–28 (2017). https://doi.org/10.1016/j.cosrev.2017.05.002
13. Timón López, C., Alamillo Domingo, I., Valero Torrijos, J.: Approaching the data protection impact assessment as a legal methodology to evaluate the degree of privacy by design achieved in technological proposals. a special reference to identity management systems. In: Proceedings of the 16th International Conference on Availability, Reliability and Security, pp. 1–9 (2021)
14. Wairimu, S., Iwaya, L.H., Fritsch, L., Lindskog, S.: On the evaluation of privacy impact assessment and privacy risk assessment methodologies: a systematic literature review. IEEE Access **12**, 19625–19650 (2024). https://doi.org/10.1109/ACCESS.2024.3360864