MAY 2025

WHITEPAPER

Guidance for DPIA practices from EU-funded projects

PAROMA-MED
Privacy Aware and Privacy Preserving Distributed and Robust Machine Learning

encrypt

TRUMPET

flute

oncovalue

HARPOCRATES

WATCHING THE RISK FACTORS
WARIFA

TITAN

CONTENTS

1. Introduction

In an increasingly data-driven world, large volumes of personal data fuel innovation, deliver services, and drive decision-making. Research and innovation across the European Union, via the funded projects under programs such as Horizon Europe, Digital Europe, or the Connecting Europe Facility involve sharing of personal data, for advanced technology development. While data processing offers substantial benefits, it also poses significant risks to the fundamental rights and freedoms of individuals. Since its enforcement in May 2018, the General Data Protection Regulation (GDPR) has aimed to address these challenges by establishing a robust framework for data protection and privacy. Rooted in the fundamental right to privacy, the GDPR establishes a robust framework for personal data protection, placing accountability and risk management at the heart of lawful processing.

It is within this context that **Data Protection Impact Assessments (DPIAs)** play a critical role. DPIAs offer a proactive approach to identifying, evaluating, and mitigating privacy risks early in the lifecycle of a project. This not only helps prevent costly violations but also reinforces accountability and fosters public trust.

This whitepaper provides practical and legal guidance on conducting DPIAs. It explores:

- **Why DPIAs are required** under Article 35 of the GDPR,
- Insights from 8 EU funded projects
- **The steps involved** in carrying out an effective DPIA—from risk identification to mitigation,
- **Best practices** drawn from these projects, recommendations, and industry experience.

By integrating privacy risk assessments into project planning and governance, DPIAs serve as a cornerstone of privacy by design and by default—as per the Article 25 of GDPR. As regulatory expectations rise and individuals become more aware of their privacy rights, DPIAs continue to be a key indicator of responsible and ethical data processing.

2. Regulatory Purpose

The DPIA constitutes a structured process for identifying and managing risks to the rights and freedoms of individuals, especially the right to the protection of personal data, caused by a specific processing operation. This is particularly the case when the processing involves the use of new technologies. The DPIA shall be conducted before the processing takes place.

In this regard, a DPIA is required when the processing operation:

- Is a systematic and extensive evaluation of personal aspects relating to natural persons, based on automated processing, including profiling, which results in decisions with legal effects for that natural person;
- Involves large numbers of special categories of personal data or personal data relating to criminal convictions;

- Constitutes a systematic monitoring of publicly accessible areas on a large scale[1].

The GDPR leaves room for the Data Protection Authorities (DPAs) to define new processing operations that they deem necessary to perform a DPIA or compile public lists with processing operations for which no DPIA is required[2].

The scope of a DPIA depends on the specific assessment being conducted. It may focus on:

(i) a single processing activity (e.g., automated deletion of outdated patient records);
(ii) a set of closely related processing operations (e.g., the entire workflow of collecting, encrypting, and analyzing patient data for research purposes), or
(iii) a broader technology or system (e.g., the deployment of an AI-driven diagnostic tool in a healthcare setting).

The individual conducting the DPIA is responsible for clearly defining the scope of the assessment and determining which processing activities should be included.

2.1 Rationale behind the DPIA

There are two main reasons for performing a DPIA. First, when it is mandatory either because of the text of the GDPR or because of the decision of a national DPA. Second, whenever a controller decides to use the DPIA for accountability reasons. A DPIA not only helps controllers comply with GDPR requirements but also demonstrates that appropriate measures have been taken. Thus, DPIA is always an option for controllers who wish to manage and enhance their compliance with applicable data protection laws. In all cases, it is good practice to review the DPIA at regular intervals based on the stage of development of the envisaged processing operations. A DPIA should be a living document that tracks evolving processing risks and adapts mitigation measures accordingly.

2.2 Main elements in a DPIA

The responsibility to carry out a DPIA lies with the controller, since it is the entity that decides the means and purposes of any processing. When a processing operation involves joint controllers, they must clearly define their respective responsibilities. Their DPIA should specify which party is accountable for implementing the measures aimed at mitigating risks and safeguarding the rights and freedoms of data subjects.[3] Each controller should communicate

[1] Article 35(3) GDPR.

[2] Article 35(4),(5) GDPR. Available guidance by DPAs can be found on the website of EDPB: Data Protection Impact Assessment (DPIA) | European Data Protection Board. The establishment of public lists of those processing operations for which a DPIA is mandatory is obligatory for the DPAs, while the publishing of the lists of the processing operations exempt from the requirement of a DPIA is optional.

[3] EDPB 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, (wp248rev.01). Available at: JUSTICE AND CONSUMERS ARTICLE 29 - Guidelines on Data Protection Impact Assessment (DPIA) (wp248rev.01)

their requirements and share relevant information while ensuring that trade secrets, intellectual property, confidential business information, or potential vulnerabilities are not compromised. Moreover, when processors are part of the processing operations, e.g., a company providing IT support, they are under the obligation to assist controllers by providing clear information on their activities. In some cases, it may be useful to seek for the views of data subjects or interested stakeholders, e.g., through a survey or a study, even though this usually comes with a high additional cost and time. In the context of a consortium in an EU-funded project multiple parties may qualify as controllers, meaning that each of those controllers, jointly or separately, may have the obligation to perform a DPIA for specific processing operations for which it decides on the purposes and the means.

The GDPR does not specify any particular methodology for the DPIA, but rather provides a list of minimum requirements to be included therein:

- A systematic description of the envisaged processing operations and the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller;
- An assessment of the necessity and proportionality of the processing operations in relation to the purposes;
- An assessment of the risks to the rights and freedoms of data subjects;
- The measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with the GDPR taking into account the rights and legitimate interests of data subjects and other persons concerned.

As long as these elements are found in the DPIA, it is up to the controller to further specify the exact form and content.

3. Key Components of a DPIA and application in each collaborating projects

3.1 PAROMA-MED

PAROMA-MED[4] stands for Privacy Aware and Privacy Preserving Distributed and Robust Machine Learning for Medical Applications. The project develops, validates and evaluates a platform – based hybrid-cloud delivery framework for privacy and security assured services and applications in federative cross-border environments. It aims at proposing a rich and secured platform to build AI and ML models and to guarantee better privacy and security capabilities in the healthcare domain.

[4] https://paroma-med.eu/

Description covering the purpose, scope, and context

The project is developing new architectures, technologies, tools and services. PAROMA-MED project approached privacy protection of personal data, including also health related records, by adhering to GDPR compliance and related processes.

The data platform is the place where the data ingested inside the PAROMA-MED ecosystem reside. Though the project uses synthetic data nevertheless since its potential use in real world in future DPIA drives the development of PAROMA-MED data platform and this is the basis for DPIA:

- Federation: this ensures that data is simultaneously part of the PAROMA-MED ecosystem but also remain local to the nodes, preventing any transmission of those sensitive data over the network.
- Security: with one separate data platform utilized for each installation of the PAROMA-MED platform in the edge nodes, this guarantees that data is not transmitted over the network.
- Privacy: GDPR has established stringent guidelines for data handling to safeguard fundamental rights, there is concern that compliance often centers on formal obligations that mainly address accountability after a data breach. In response, we propose an innovative approach, termed "Functional GDPR," emphasizing full transparency and control over data flows and continuous verification of entities involved in transactions.
- Usability: as the data is spread between different edge nodes, this allows the data is discoverable in a structured and standardised way. This includes also how data is generated, advertised, discovered and utilised (data life cycle).
- Traceability: this ensures that data and models are traceable in case of leaks (zero risk does not exist).

Assessment of necessity and proportionality

The PAROMA-MED project is primarily about the protection of personal data, the aspect of security and privacy is fundamental to the success of the project.

PAROMA-MED platform involves processing of synthetic data (as per DoW) using new technologies, an assessment is being carried out to identify the impact of the related operations on the protection of personal data. PAROMA-MED project and the consortium is carrying the work to provide a proof-of-concept approach (TRL-4) subject to further evaluation and potential improvement before applied to production environment, a DPIA is performed to ease the adoption and maximize any potential uptake of the outputs.

Data Protection Risks

The fact that the project builds on an approach with the least or even zero data transfer is considered as a guarantee regarding risk minimisation. However, the following risks are identified.

IDENTIFIED RISK	DESCRIPTION
DATA BREACH DUE FROM CODE TO DATA	Code that is sent to data maybe malicious and leak information to third party servers
INFERENCE ATTACK	AI model can be used from third parties to reveal information on participants and data sets used
ABSENCE OF USER CONSENT	User has not provided consent, so usage of the data might be out of concrete constraints
REVOKE OF CONSENT	After use of data, user may remove or modify consent
WM TRACEABILITY	Images without WM may leak out

The project aims at delivering transparency by allowing complete control of the consent options as well as educated and well-informed decisions.

The foreseen attestation mechanisms are subject to research and adoption challenges for the establishment of a more efficient handling of GDRP legislation that promotes both business and solution growth without any sacrifice with respect to privacy protection.

The DPIA serves as a living document and supports in documentation of the project's compliance efforts.

3.2 TRUMPET:

Description covering the Purpose, scope, and context.

The TRUstworthy Multi-site Privacy Enhancing Technologies (TRUMPET[5]) project aims to research and develop novel privacy enhancement methods for federated learning (FL), and deliver a highly scalable federated artificial intelligence (AI) service platform for researchers, enabling AI-powered studies of siloed, multi-site, cross-domain, cross-border European datasets with privacy guarantees that exceed the requirements of the GDPR. Another important objective is the research, development and promotion of a novel privacy metric and tools for the supporting GDPR compliance of FL implementations.

The TRUMPET platform will be piloted and validated in the following use cases:

- Use Case 1: Integration of multiple markers for immunotherapy response prediction in metastatic non-small cell lung cancer
- Use Case 2: Late side effects of intensity-modulated radiotherapy in head & neck cancer patients
- Use Case 3: Eligibility criteria for stereotactic body radiotherapy

[5] https://trumpetproject.eu/

These use cases will evaluate the feasibility and privacy risk of the TRUMPET platform and demonstrate how the platform enables several hospitals to collaboratively build a machine learning (ML) model while preserving the privacy and confidentiality of patient data.

Assessment of necessity and proportionality:

The primary goal of processing activities in the TRUMPET project is to validate the privacy-enhancing technologies (PETs) incorporated TRUMPET platform. Moreover, with the use of the platform, the researchers develop statistical and AI models without transferring data out of the hospital infrastructure (data holder node) in a privacy preserving way. Given the above, the use of retrospective medical data in the project is necessary to conduct the defined use cases. The scope of the data is clearly defined in a study protocol and the categories of data used were kept in proportion to the scope of use cases. Before making use of medical data, partners made sure to minimize the processed data through anonymization or pseudonymization, where full anonymization is not possible. Finally, TRUMPET provides privacy metrics to quantitatively assess the identification privacy risks assumed by data holder nodes when exchanging data through the platform.

Which kind of data is being processed and why (what are the requirements?)

Patient data used in TRUMPET project are sensitive and relevant to health aspects, such as information on cancer diagnosis, radiotherapy, smoking addiction, gene alterations, stage of therapy etc. Such a multidimensional approach allows to obtain a comprehensive overview of patients' health condition and train the AI system to come up with tailored diagnosis using a multitude of indicators.

Risk assessment: Identification and evaluation of risks to data subjects.

Conducting a DPIA for the TRUMPET platform is recommended due to novel technologies (FL) that the project involves, such as FL, as well as due to processing of sensitive medical (retrospective) data of patients.

In the risk assessment, the TRUMPET DPIA relies on the methodology proposed by the French DPA (CNIL), where the risk is calculated based on its severity and likelihood of occurrence. Four levels of risk are possible stemming from the combination of the aforementioned parameters: negligible, limited, significant, and maximum.

In the DPIA risks are addressed not per use case, but rather generally based on the type of the impact they cause to data, i.e., illegal access, unauthorized modification or loss of data, and their source, i.e., whether they originate from internal (e.g., a staff member) or external (e.g., a third-party attacker) actors.

Mitigation measures: Safeguards and controls to address identified risks.

The core principle applied to all medical data within the TRUMPET platform is risk mitigation through data minimization. This is achieved by applying data de-identification techniques, specifically pseudonymisation or anonymisation, to reduce the potential impact on data

subjects. In addition, data processing is conducted in a federated manner, ensuring that personal data remains within local environments while still supporting collaborative research.

In particular, the individual level data itself remains in the data holder node and is pseudonymized (i.e., directly identifying features are removed) or anonymized. If a query is performed, the following steps are performed one or more times:

1. The local data of the data owners is aggregated (e.g., an average of a given function over this data is computed, a sample of it is taken and an aggregate is computed over it).

2. The partial results are encrypted and/or privatized in such a way that without the collaboration of the data holder, functions which are computed from these partial results can't reveal any sensitive information.

3. The partial results of the data holders are aggregated into a global result. This global result can then be the answer to the query or just one step in the computation (e.g., a gradient update to a model). If encryption is used, at some point after the aggregation and before outputting the answer to the query, information is decrypted with the collaboration of the data holders, who can at that time also verify that all steps have been performed correctly (even if they can't see the intermediate results in order to protect the data of other data holders). As a result, the processing of the data in TRUMPET platform contains multiple layers of protection of the data and ensures that no sensitive information can be extracted from the data in the data holder nodes.

The DPIA contains a section dedicated to the applicable technical and organisational measures per data holder. Some of the further measures implemented in the project are: storage of data in the data holder nodes with prior checks of the quality of data and application of logical access controls to ensure the integrity of data.

Documentation and reporting: Recording the DPIA process and findings.

For the TRUMPET project, the DPIA follows a two-step approach, based on the Privacy Impact Assessment (PIA) methodology recommended by CNIL[6], consisting of a descriptive and an appreciative part. The descriptive part provides a detailed overview of the processing activities at hand, in terms of nature, scope, context, and purposes. The appreciative part focuses on an analysis of fundamental principles of data protection law, and of how data subjects can exercise their data subject rights in relation to the processing activities in scope of this DPIA.

Approach taken by the project

The DPIA follows a detailed approach that describes thoroughly the architecture of TRUMPET platform, identifies the crucial data processing operations, and the categories of data

[6] Available in English: https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-1-en-methodology.pdf. The methodology is further supported by a catalogue of controls aimed at complying with the legal requirements and treating the risks, and indicative examples. Available in English: PIA, knowledge bases.

processed and mapping potential risks, listing listed the applicable technical and organisational measures to mitigate them.

3.3 FLUTE:

Description covering the Purpose, scope, and context.

The Federate Learning and mUlti-party computation Techniques for prostatE cancer (FLUTE[7]) project aims to advance and scale up data-driven healthcare by developing novel methods for privacy-preserving cross-border utilization of data hubs.

The FLUTE platform will be piloted and validated in the use case regarding robust prostate cancer prediction using FL across borders of four European countries.

The data used in the context of the use case stems directly from medical centres located in three European countries, and indirectly from a partner that enlarges and geographically diversifies FLUTE's datasets by providing retrospective data from proprietary data collections of clinics with which they collaborate.

The FLUTE platform will provide the privacy guarantees and scalability enabling deployment of such multi center, cross-border healthcare AI models. Overall, the main expected results of the project are:

- Development and validation of a cross-border federated AI solution on the FLUTE platform for the diagnosis of csPCa.
- Novel intermediating FLUTE platform that will make the siloed datasets available to researchers under privacy protection that complies with GDPR requirements, in standard FL settings and AI model development workflows.
- Validation of the Barcelona Risk Calculator models (BCN-RC 1 and 2) with the inclusion of imaging biomarkers extracted from mpMRI/bpMRI for the detection of csPCa with data from multiple regions. The Barcelona models constitute statistical models to determine csPCa using a cohort of patients and variables such as age at the time of biopsy, PCa family history, etc.
- Generation of synthetic clinical data and images to train convolutional neural network models in a privacy-preserving way, i.e. without the risk of re-identifying sensitive data of patients.

Assessment of necessity and proportionality:

The primary goal of processing activities in the FLUTE platform is to validate the Barcelona Risk Calculator models (BCN-RC 1 and 2) with the inclusion of imaging biomarkers extracted from mpMRI/bpMRI for the detection of csPCa with data from multiple regions. The secondary goal is to achieve robust performance in csPCa cancer detection, considering variations in the prevalence of the disease in different regions of Europe, and to develop and validate a cross-

[7] https://www.fluteproject.eu/

border federated AI solution on the FLUTE platform for the diagnosis of csPCa in different regions of Europe.

Which kind of data is being processed and why (what are the requirements?)

Patient data used in FLUTE project are sensitive and relate to health, such as imaging data from prostate MRI, clinical variables from laboratory findings, imaging biomarkers, etc. Using diverse data categories allows to obtain a comprehensive overview of patients' health condition and to train the AI system in a federated way to come up with tailored diagnosis using a multitude of indicators. Moreover, the project deploys synthetic data in the form of 1-D and 2-D objects.

Risk assessment: Identification and evaluation of risks to data subjects.

Conducting a DPIA in FLUTE project is recommended due to novel technologies (FL) that the project involves, as well as the processing of sensitive medical (retrospective) data of patients.

Similarly to the TRUMPET project, the CNIL methodology is applied for the categorization and identification of risks. Hence, the main risks relate to illegal disclosure, unauthorized modification or loss of the data stored in the data nodes and used through the FLUTE platform.

However, the DPIA needs to take into account potential additional risks related to the generation and use of synthetic data for AI model training and a complex data flow, which includes both the FL platform, and a central storage used for processing of certain image data.

Mitigation measures: Safeguards and controls to address identified risks.

The approach described in the context of the TRUMPET project is applicable also in the FLUTE project, i.e. the main mitigation measure for medical data is data minimization by means of pseudonymisation or anonymisation to limit the impact on data subjects. The potential of such a mitigation measure is enhanced because of the federated processing of data that takes place in the FLUTE platform, as described in the following paragraphs.

The main principle in the FLUTE project is that no patient data leaves the data holder premises (data nodes). The data holder node via FLUTE Platform exchanges encrypted messages with other data holders to collaboratively compute aggregates in such a way that under the security assumptions:

- no sensitive information can be revealed from these exchanged messages;
- only a privatized version of the aggregate/model/statistic/etc. the data holder agreed in advance to compute can be revealed. Researchers will not see or have any direct access to the data when it is stored in the data holder node. They will interact with it exclusively through the FLUTE platform by making queries.

The DPIA contains a section dedicated to the applicable technical and organisational measures per data holder. Some of the further measures implemented in the project are: partitioning of data to avoid the crosslink of cases/patients and also robust organizational measures, such as allocation of tasks under a joint controllership arrangement, internal policies, including, for

instance the prohibition of storing data on user and removable devices. The implemented measures will be tested by external pen-testers.

Documentation and reporting: Recording the DPIA process and findings.

Questionnaires were disseminated to data holders to identify data flows and draft a data mapping document. Next, the data flow was discussed with technical partners and all of the consortium. Based on this, DPIA draft was prepared and circulated to other partners for input and validation.

Approach taken by the project

FLUTE constitutes a pioneering research endeavour that promotes robust prostate cancer prediction using FL in a privacy-preserving way. The use of synthetic data adds to the complexity of the project. Thus, in order to support a privacy by design approach, preparation of the DPIA is imperative to understand the architecture of the platform, identify the crucial data processing operations and the categories of data processed, observe the applicable technical and organisational measures to mitigate existing and potential risks, and compile a list of action points that committing the partners to further improve the compliance measures. This approach reaffirms the function of the DPIA as a living document that shall be subject to regular reviews according to the project's work progress.

3.4 ENCRYPT

Description covering the Purpose, scope, and context

Initiated in June 2022, ENCRYPT[8] is an innovative collaborative project, designed to integrate privacy-preserving technologies, including Fully Homomorphic Encryption (FHE), Trusted Execution Environments (TEE), and Differential Privacy (DP), across three key sectors. The project aims to revolutionize the processing and protection of sensitive data in the digital age.

The project´s, Data Protection Impact Assessments (DPIAs) were conducted for three specific use cases to evaluate the implications of implementing the ENCRYPT platform gathered through structured questionnaires. Their scope was limited to processing activities conducted during the testing phase of the ENCRYPT platform in the relevant use cases. Each use case partner conducted pre-DPIAs at an early stage, allowing for discussions and feedback during the consortium meeting held in June 2023. Subsequently, comprehensive DPIAs were developed and completed.

Assessment of necessity and proportionality

The assessment of necessity and proportionality played a fundamental role in the DPIAs for the ENCRYPT use cases. The project is specifically designed to facilitate the circulation of both personal and non-personal data while ensuring robust privacy protections, particularly in

[8] https://encrypt-project.eu/about/

fields such as cooperative oncology, where multiple entities require access to the same personal data.

In some cases, project partners developed synthetic datasets consisting of artificial data. However, in instances where actual data was utilized, an assessment was conducted to determine whether the data had merely been pseudonymized or fully anonymized, in accordance with Recital 26 of the GDPR, to ensure that there was no risk of re-identification.

Risk assessment: Identification and evaluation of risks to data subjects

In the context of the first use case of the project, which pertains to the financial technology (FINTECH) domain, two sub-cases were examined. The first involved an assessment of data privacy through the ENCRYPT recommendation engine, where privacy-preserving technologies could be adopted based on the assessment results. The second sub-case involved the pseudonymization of data before its transfer from a bank to a partner, where it was used to train AI models for a software suite and the ENCRYPT recommendation engine. In both instances, the implementation of privacy-preserving technologies or pseudonymization constituted technical measures that significantly reduced the level of risk and mitigated the impact of processing activities on data subjects' rights and freedoms. The DPIA identified six risks, all of which were classified as low risk due to their low likelihood of occurrence and low severity of consequences (namely: : processing of sensitive data or data of highly personal nature; systematic monitoring; prevention of the data subject from exercising their rights related to data protection due to processing; discrimination/stigmatisation; damage to the reputation; unauthorised or unjustifiable access, transfer, sharing or publishing of data).

The second use case concerns the medical domain and seeks to evaluate the ability of the ENCRYPT technical solution to safeguard sensitive personal information without unduly restricting its circulation. The data processed in the testing phase consisted of imaging medical results, while text data was limited to synthetic data only. No personal data was transferred to third parties, including external medical practitioners. To ensure compliance with the GDPR, patient health, genetic, and behavioral data, along with personal identifiers, were replaced with ‘fake data’. However, because DICOM medical images constitute personal data and cannot be anonymized, pseudonymization techniques were applied. Although the processing of health-related data typically involves high-risk considerations, the DPIA determined that the overall risk level ranged from low to medium.

In the third use case (so called CTI domain), some partners acting as data providers provide data and logs from their systems for the extraction and enrichment of the Cyber Threat Information (CTI) in a secure and privacy-preserving manner through another partner, which will provide the necessary tools for the extraction of CTI from the data and the correlation and enrichment of all the CTI extracted by combining the different data of the data providers and by leverage information from external sources when possible. These tools will be adjusted to work with the ENCRYPT privacy-preserving framework. No special categories of personal data (Art. 9(1) GDPR) nor data relating to criminal convictions (Art. 10 GDPR) are foreseen to be collected. The risk of processing of sensitive data or data of highly personal nature is low and identified as such because of the technical setting of the CTI platform, which should prevent special categories of data from being processed in the testing phase.

Mitigation measures: Safeguards and controls to address identified risks

A distinctive aspect of the ENCRYPT project lies with the fact that it aims at implementing PPTs to facilitate data sharing within federated environments. The fact that the processing of personal data (if any) occurring in the testing phase is meant to facilitate the development and application of privacy-preserving technologies does not per se reduce the level of risk but significantly contributes to limiting the probability that high risks emerge.

1. In the first use case (FINTECH domain), the peculiar modalities for the processing of data help reducing the overall level of risk.

2. In the second use case (MEDICAL domain) all the identified risks may present *per se* high severity of impact; however, the fact that the personal data processed in the testing phase only consist in images (relating to hundreds of patients) significantly mitigates the consequences of a possible risk. Also, the level of risk in terms of probability is reduced because of the high quality of the technical infrastructure developed and used by the technology owner and the limited circulation of the personal data in question. In any case, it goes without saying that in the use case at hand the images (relating to brain tumour diagnosis) cannot be subject to further technical processing that would lead to 'anonymise' or 'pseudonymise' them.

3. In the third use case (CTI domain) most of the more likely or impactful risks seem to be properly mitigated, by virtue of the adoption of adequate technical measures.

Documentation and reporting: Recording the DPIA process and findings

The ENCRYPT project adopted a DPIA methodology based on the template developed by the Vrije Universiteit Brussel's DPIA Lab. This methodology is tailored to GDPR requirements and incorporates guidance from European Data Protection Authorities

To support risk assessment, the project recommended the use of the Spanish Data Protection Authority's Risk Assessment Tool, a resource that provides a structured framework for evaluating potential risks associated with data processing.

The DPIA lab scalable template, is used for a reasoned and comprehensive assessment on the characteristics of the processing activities, their legal implications, the level of risks, the applicable technical and organisational measures. The role of the various partners involved as data controllers rather than processors, in order to determine the applicable obligations and requirements and to ascertain compliance with the privacy-by-design and privacy-by-default principles, including the definition of a proper legal basis for the processing activities.

3.5 HARPOCRATES

Description covering the Purpose, scope, and context

The HARPOCRATES project (2022–2025) aims to develop innovative cryptographic schemes that allow for the federated processing of data while preserving privacy. The primary objectives include the creation of digitally blind evaluation systems designed to eliminate

biases related attributes such as geography, gender, and race. The project also focuses on enabling the classification of encrypted data without the need for direct decryption. HARPOCRATES utilizes advanced cryptographic techniques, such as Functional Encryption (FE) and Hybrid Homomorphic Encryption (HHE), to create privacy-preserving machine learning models. These models can classify encrypted data and make accurate predictions directly on ciphertexts across federated data spaces. Given the sensitive nature of the data involved in this research, which includes health and cybersecurity information, a DPIA is essential. This assessment helps identify potential risks and ensure that strong data protection measures are implemented for federated learning and privacy-preserving machine learning scenarios.

Assessment of necessity and proportionality

HARPOCRATES operates in compliance with the GDPR, specifically under provisions that relate to scientific research (Article 9(2)(j)) and public interest (Article 6(1)(e)). The project ensures that personal data is processed only when necessary for the development of HARPOCRATES cryptographic technologies, while adhering to principles of reliability and accountability. The project has adopted a privacy-by-design framework after evaluating alternative approaches, which were deemed inadequate in balancing the security and efficiency of encryption-based solutions.

This strategy includes a combination of protected public datasets, anonymization techniques for personal identifiers, and privacy-preserving machine learning methods to maintain stringent data protection standards. Data collection is strictly limited to predefined objectives, including validating encryption methods for medical research related to sleep disorder treatments and cybersecurity applications.

Which kind of data is being processed and why?

The HARPOCRATES project collects and processes a range of datasets primarily technological development purposes. The following datasets are included in the project:

DATASET	CHARACTERISTICS	PURPOSE IN HARPOCRATES	CONTAINS PERSONAL DATA	ANONYMIZED BY HARPOCRATES
MNIST	Handwritten digit images; used for machine learning benchmark tests; non-personally identifiable.	Benchmarking encryption techniques for machine learning model performance.	No	N/A
MIT-BIH ARRHYTHMIA	ECG signals from real individuals; used for testing encryption techniques.	Testing encryption methods for biomedical signal processing in health applications.	No	N/A
SIESTA	Sleep recordings (590 subjects); includes healthy	Validating privacy-preserving methods	Yes, special category (age, sex, height,	Yes

	subjects and those with sleep disorders; bio-signals and respiratory events data.	for sleep medicine research.	weight, blood pressure, heart rate, sleep disorder diagnosis)	
ESADA	Polysomnography and cardiorespiratory polygraphy recordings (400 total); collected from European sleep centers.	Testing privacy-preserving encryption techniques for sleep disorder analysis.	Yes, special category	Yes
SARGA **(SPAIN)** VENETO **(ITALY)**	Behavioral data from employees; includes names, surnames, and IP addresses. Time period & host/subnet identifiers. Pre-processed log files for cybersecurity threat analysis. Windows Event Log Dataset. Collected from volunteer users, includes network security data.	Enhancing cybersecurity threat detection and intelligence sharing. Testing and developing federated learning models for cybersecurity threat detection. Enhancing AI models for anomaly detection and cybersecurity threat intelligence.	Yes	Yes

Personal data is collected from specific datasets where necessary, but only **four datasets** (SIESTA, ESADA, SARGA, and VENETO) contain personal identifiers, which are **anonymized** once gathered by HARPOCRATES. The purpose of processing this data is to develop secure cryptographic methods that will enhance privacy and security in medical and cybersecurity applications.

Risk assessment: identification and evaluation of risks to data subjects

A threshold assessment was conducted to evaluate whether the HARPOCRATES project involves high-risk data processing activities under the GDPR, based on guidance from the European Data Protection Board. The assessment identified moderate risk and three clear high-risk indicators: Processing of sensitive health data, inclusion of vulnerable individual's data, and Use of emerging technologies (e.g., functional encryption, federated learning). The DPIA identified and evaluated several potential risks related to the handling of sensitive data within the project:

- **Unauthorized access**: There is a risk of data breaches that could lead to unauthorized access, exposure, or destruction of sensitive health or employee data. This

underscores the need for robust security measures, including encryption and access control.

- **Re-identification of (Pseudo)anonymized Data**: Although anonymization and pseudonymization techniques are employed, there remains a residual risk of re-identification, particularly if encryption keys are compromised or datasets are linked with external sources.
- **Function creep**: There is a potential for data to be used beyond its original, intended purpose (e.g., repurposing medical or cybersecurity data), raising ethical and legal concerns. Strict purpose limitation is necessary to mitigate this risk.
- **Data integrity and performance issues**: The implementation of encryption methods to protect sensitive data may lead to performance bottlenecks or errors in data processing, which could affect the accuracy of machine learning models used in medical or cybersecurity contexts.

To mitigate the identified risks, HARPOCRATES implements robust security safeguards, including strong encryption, access controls, and secure key management practices.

Documentation and reporting: recording the DPIA process and findings

The DPIA process is documented in the project DMP and D6.1/2 (Research ethics report), including detailed records of data flows, risk assessments, and the security measures that have been implemented. The HARPOCRATES project ensures full transparency by maintaining compliance records and regularly reviewing its adherence to privacy standards. The findings from the DPIA are evaluated and updated as needed to ensure that privacy-by-design principles are followed throughout the project lifecycle. Regular data protection analyses are conducted to verify that appropriate safeguards are in place, and data-sharing agreements are enforced to prevent the misuse of sensitive data.

3.6 ONCOVALUE

Description covering the Purpose, scope, and context

The ONCOVALUE[9] project brings together cancer institutes and commercial businesses, working to implement value-based oncology care by enabling and guiding cancer clinics to collect, harmonize and analyse high-quality Real-World Data (RWD) in real-time. The aim is to unlock the full potential of digital sources for RWD collected in European cancer hospitals to ease health regulatory and health technology assessment (HTA) decision-making on cost-effectiveness of novel cancer therapies. To achieve this, the consortium builds up data collection and processing capabilities of leading European cancer hospitals to create high-quality clinical, quality of life, and adverse events data-sources. With the use of powerful AI technologies, ONCOVALUE transforms unstructured data originating from medical notes and medical images into structured data to enable analytics and Real World Evidence (RWE).

[9] https://oncovalue.org/

Assessment of necessity and proportionality

The consortium uses an existing set of instructions for a DPIA (from its partners acting as controllers) including information on what should be included in an impact assessment. The documents describe that a data controller needs to carry out a thorough assessment of the risks associated with the processing of personal data whenever a new system, software, functionality or equipment is introduced in which personal data are processed as mandated by Article 36 of the EU GDPR.

AI tools for the automatic extraction of structured outcomes from medical notes and images are developed within the ONCOVALUE project. In the context of this whitepaper, the practises followed in HUS Helsinki University hospital, the coordinator of the consortium, are described. The objective concerning medical notes is to exploit machine learning (ML) to pick information from unstructured text data and convert it to structural form by developing a Natural Language Processing (NLP) model. During this collaboration, pseudonymized personal health data is processed and shared with the research team.

However, this activity is carried out in a secure processing environment to ensure compliance with the Finnish Act on the Secondary Use of Health and Social data. The data cannot be transferred outside of the secure environment without a separate anonymisation check conducted by the local Social and Health Data Permit Authority (Findata), leaving risk for privacy concerns to a minimum. Furthermore, a separate data agreement regarding the processing of the data has been established, and all employees with access are bound by non-disclosure agreements. Additionally, the hospital's Research leadership team has granted permission to complete this research work.

Similarly, the development of the image AI model also takes part in a secure environment, and similar agreements have been acquired for its completion. The image AI model has the primary objective of enabling the automatic estimation of metastatic cancer progression from whole-body computer tomography (CT) scans, specifically focusing on liver and lung metastasis. The ultimate aim is to empower hospitals to assess which therapies are effective in the treatment of metastatic breast cancer by providing information on disease progression. The NLP tool aims to extract structured outcomes for specific variables from unstructured medical notes to enable analytics of RWE on the effectiveness of cancer therapies.

Which kind of data is being processed and why (what are the requirements?)

ONCOVALUE focuses on providing data on treatment of metastatic cancer in different cancer types such as breast cancer and non-small cell lung cancer (NSCLC). Reaching the objectives of the project requires data on patients, diagnostics, therapies, treatment costs and patient-reported outcomes (PROMs), which are analysed and processed in the context of scientific research. Data sharing between consortium partners is primarily planned to be conducted in an aggregated format but also anonymized or pseudomised data may be shared. In such cases, the partners sharing data ensure that it is conducted in a GDPR compliant manner.

The ONCOVALUE project is also developing a federated data platform (HTA Dashboard), which is envisaged to enable connecting project partners OMOP harmonised (health) data sources for the purposes of HTA. During project lifetime, the technical feasibility of the platform is

tested by connecting only synthetic data to this dashboard. Additionally, testing with real patient data is conducted in a secure environment with a limited dataset.

In conclusion, even though personal data is processed within the project, it is done in compliance with all legal requirements of the countries participating in it. The primary notion is that all development work conducted within the project fall within the scope of research, not in product/medical device development. Due to these points, the project has assessed that a DPIA is not necessary at this time.

3.7 WARIFA

Description covering the Purpose, Scope, and Context

The WARIFA[10] project ("Watching the Risk Factors: Artificial Intelligence and the Personalized Prevention and Management of Chronic Conditions") is a Horizon 2020 Research and Innovation Action focused on empowering citizens through personalized recommendations targeting modifiable risk factors. The project addresses the prevention of chronic conditions such as melanoma, cardiovascular diseases, and diabetes. As part of the action, the WARIFA app was developed to deliver lifestyle-related advice tailored to each individual's data. A proof-of-concept of the app was piloted in a real-world setting over a 12-week period.

When it comes to data protection, WARIFA followed a structured and risk-based approach that reflects the principles of the GDPR. The DPIA process began with an initial risk assessment, which was conducted in the first six months of the project and updated as the project progressed, responding to changes in the technical designs and data flows. The project is coordinated by the University Hospital of North Norway (UNN), through the Norwegian Centre for E-health Research (NSE), which also serves as the data controller. The Chief Information Security Officer (CISO) at UNN had a key role in identifying and evaluating the risks tied to personal data processing. For the DPIA itself, we used the template recommended by the Norwegian Health Directorate, which provides a practical guide for identifying and documenting privacy risks in digital health systems (https://www.helsedirektoratet.no/veiledere/personvernkonsekvensvurdering-dpia-mal)

Assessment of necessity and proportionality

The assessment covered the full WARIFA system—from data collection via wearable devices and questionnaires to how data is stored, processed using AI, and then used to give feedback to users through the app. Because the system generates tailored recommendations and early warnings about chronic disease risks, it naturally requires access to personal and sensitive information—such as demographics, health status, behavioral patterns, and even contextual data from the environment.

To evaluate necessity and proportionality, we relied on the scientific goals and statistical power calculations set out in our study protocol. Only the data needed for the AI models to

[10] https://www.warifa.eu/

work properly was collected. All data is pseudonymized and encrypted, both when stored and during transmission. Access is strictly limited to selected members of the consortium, and always under tight security controls.

Which kind of data is being processed and why (what are the requirements?)

Potential risks to users' rights and freedoms were identified using the Whatif tool. These risks were then carefully reviewed with support from the CISO. Because AI-based research often involves long-term data retention, we paid particular attention to balancing research needs with data subjects' rights. Mitigation measures include managing consent effectively, communicating user rights clearly, limiting how long data is kept, logging access to data, and following privacy-by-design practices like data minimization and secure infrastructure.

The DPIA was versioned across the lifespan of the project. The final version (v2.0) was submitted to our Data Protection Officer (DPO) in early 2024. This final version includes the complete DPIA template, comments and contributions from both the CISO and DPO, and a full record of all data processing activities.

Importantly, all personal data in WARIFA is processed based on explicit consent, in line with Articles 6(1)(a) and 9(2)(a) of the GDPR. There are no transfers of personal data outside the EU. Throughout the project, we have tried to embed the principles of lawfulness, fairness, transparency, and data minimization directly into the system's design and daily operations.

3.8 TITAN

Description covering the Purpose, Scope, and Context

TITAN[11] project operates within the evolving paradigm of European Data, ecosystems designed to facilitate the secure exchange of data among multiple organizations. However, these spaces present privacy and security challenges, which have driven the development of advanced mechanisms such as Privacy-Preserving Enablers (PPEs). In this context, TITAN aims to develop secure and reliable capabilities for confidential data processing and sharing within the European Open Science Cloud (EOSC) ecosystem. TITAN's open-source platform ensures privacy preservation and compliance with EU regulatory and legal frameworks, with a strong emphasis on AI and secure data processing.

Assessment of Necessity and Proportionality

Since TITAN processes sensitive data and employs advanced technologies such as AI and Federated Learning, conducting a Data Protection Impact Assessment (DPIA) is essential to:

- Identify potential risks in the processing of sensitive data.
- Evaluate the safeguards implemented to mitigate such risks.

[11] https://titan-eosc.eu/

- Ensure compliance with the General Data Protection Regulation (GDPR) and other European regulatory frameworks.

For data sharing, TITAN leverages a FIWARE Data Space architecture, incorporating Privacy-Preserving Enablers (PPEs) and a Policy Enforcement Point - Policy Decision Point (PEP-PDP) framework to enforce contextual, real-time access control. This ensures GDPR-compliant data usage based on explicit purpose limitation and traceability.

The PDP-PEP system enables:

- Centralized definition of privacy policies, ensuring uniform access rules.
- Real-time enforcement to block unauthorized access.
- Decision logging and auditing to meet the GDPR transparency requirements (Art. 5.2).
- Dynamic access control, adapting to different scenarios and privacy requirements.

This approach allows TITAN to comply with key GDPR principles, such as data minimization (Art. 5.1.c) and purpose limitation (Art. 5.1.b), while also facilitating the protection of data subject rights (Art. 15-22).

TITAN follows the International Data Spaces (IDS) recommendations, which provide technical measures (encryption, access control) but also require organizations to implement organizational measures (policies, consent management) to comply with GDPR.

Which kind of data is being processed and why

TITAN categorizes its datasets into two main groups: Agricultural Datasets and eHealth Datasets, each serving distinct research and technological purposes.

- Agricultural Datasets focus on meteorological, phenological, and satellite data to support predictive modeling for crop management, disease forecasting, and environmental monitoring. These datasets primarily originate from Spain and Italy, providing valuable insights for precision agriculture and sustainability.

Dataset	Description	Partner	Country
RED FARA	Observations on crop phenology and diseases, meteorological data.	SARGA-ITA	Spain
SIAR NETWORK	Meteorological data from stations in irrigated areas of Aragón.	SARGA-ITA	Spain
AEMET	Meteorological data and forecasts across Spain.	SARGA-ITA	Spain
DISEASES AND PHENOLOGY PREDICTIONS	Predictive model results based on meteorological and phenological data.	SARGA-ITA	Spain
COPERNICUS	Results of running the prediction models with the meteorological and phenological datasets of the fara network	SARGA-ITA	Spain

MORGANA	Observations on crop phenology and diseases, meteorological data.	Veneto	Italy
METEO IDRO NIVO	Meteorological data from stations in irrigated areas of Aragon.	Veneto	Italy
AGROMETEO	Processed agrometeorological data for farmers and technicians.	Veneto	Italy
COPERNICUS (VENETO)	Multispectral images and satellite meteorological data.	Veneto	Italy

- **eHealth Datasets** are centered on sleep studies and apnea research, contributing to advancements in medical diagnostics and machine learning applications. Sourced from leading research institutions in Germany, France, and Finland, these datasets adhere to strict anonymization protocols to protect patient privacy.

Dataset	Description	Partner	Country
ESADA	Polysomnography and polygraphy records of sleep apnea patients (2007-2021).	Charité	Germany
PHENOSLIP	Clinical polysomnography data for pediatric neurodevelopment and sleep studies.	INSERM	France
SMARTSLEEP LAB DATA	Polysomnography records from patients with sleep apnea and healthy volunteers (2022-2024).	UEF	Finland

Risk assessment

The Data Protection Impact Assessment (DPIA) evaluates potential risks associated with managing sensitive data in the TITAN project, which involves health and cybersecurity information, as well as advanced technologies like AI and federated learning.

Critical risks include unauthorized data access, Security breaches or improper access may expose sensitive information, affecting system integrity.

Another critical issue is function creep, which can lead to ethical and legal implications. Data collected for a specific purpose, such as cybersecurity threat detection, might inadvertently be repurposed for unintended uses. To prevent this, the purpose limitation principle (Article 5.1.b of the GDPR) must be applied through well-defined usage policies and regular audits.

Additionally, infrastructure attacks pose a threat to system availability and reliability. These include unauthorized access, security failures, and errors in data processing. To address these vulnerabilities, it is crucial to implement advanced protection mechanisms and continuous security audits to ensure a secure digital environment.

TITAN addresses these risks by aligning with IDS recommendations and GDPR requirements, ensuring transparency (Art. 5.2), purpose limitation (Art. 5.1.b), and robust breach notification protocols (Art. 33–34). The PDP-PEP framework further enforces dynamic access controls and audit trails, reducing risks while enabling secure, collaborative AI workflows.

4. Best Practices derived from projects

4.1 PAROMA-MED

The following best practices have been identified from PAROMA-MED project:

The first activity for the project was to educate on GDPR in the early stages of the project. This was done by organizing GDPR training to all the partners to enable the understanding of the requirements and obligations established under EU privacy protection law. This helped the consortium participants, understand this crucial framework for protecting personal data in the EU to ensure compliance. As PAROMA-MED aims to ensuring Privacy by Design & Default this was particularly helpful to map how the platform was built with data protection measures, apply data minimization and safe guarding data through encryption, anonymization and pseudonymization.

There were regular discussions via the monthly meeting and plenary sessions that ensured- data processing activity has a lawful basis (e.g., consent, legitimate interest) and the interfaces/connectors developed comply with GDPR's strict consent requirements.

Workshop with partners was organized for DPIA which continued into several exchanges to understand the main flows and processes involved and at each stage the functionalities and data storage and format was considered. The risks and their mitigation measures were discussed and applied.

4.2 TRUMPET & FLUTE

In both projects one of the best practices identified is to build a close and uninterrupted collaboration with technical partners already from the very early stages of the project. This is accomplished in several ways, including by organising workshops to understand the particularities of different key legal concepts. This approach ensures that legal requirements are translated into concrete guidelines on technical processes. For DPIA in particular, the drafting of the first iteration required regular exchanges with technical partners, especially for describing precisely the main functionalities of the platforms, according to the current stage of development, mapping of the key risks, and defining appropriate mitigation measures. These exchanges were facilitated by the prior organisation of dedicated workshops and regular updates of input is expected to continue until the final version of the DPIA.

Moreover, the legal partner in both projects closely followed the meetings of the technical work packages to identify potential legal issues as they arose during platform development and to be able to provide timely and appropriate guidance. To support the execution of complex legal tasks, such as the DPIA, data owners were asked to identify and describe the categories of data collected within the first few months of each project.

This then led to the mapping of the data flows involved in the development of the platforms. Once the platform architecture became clearer, the legal partner prepared a data sharing agreement for the consortium. The draft was widely consulted with other partners. As a result, the agreement aimed to clarify each partner's role in relation to data processing, define the

scope of the data to be used, and set out the applicable pseudonymization or anonymization measures.

Lastly, the inclusion of specific action points in each iteration of the DPIA makes next steps actionable, commits partners to their actual involvement, and ensures that the DPIA reflects the necessary legal requirements for the respective level of projects' progress.

4.3 ENCRYPT

The following best practices have been identified from conducting DPIAs for the ENCRYPT project:

Firstly, for projects involving a high degree of technical complexity, it is essential to provide proactive support to partners and experts who may not have a legal background. This ensures they can navigate the requirements and obligations established under EU data protection law.

Another best practice is the active involvement of partners' legal counsels and Data Protection Officers (DPOs) in both validating DPIA outcomes and, where feasible, participating in earlier discussions regarding the assessment's content. While DPIAs are crucial for project compliance, the expertise of these professionals—who are well-acquainted with each partner's business context—enhances the effectiveness of the assessment. Their engagement also helps to develop DPIAs as scalable tools that can support internal compliance efforts beyond the scope of the project.

Finally, for projects with a large number of partners, it is advisable to establish structured procedures to ensure the regular exchange of information on factors that may necessitate a reassessment of DPIA outcomes. In the case of ENCRYPT, questionnaires were used to systematically collect relevant details about processing activities. These questionnaires were disseminated annually to facilitate the review and, where necessary, the amendment of DPIAs and their conclusions.

Maintaining up-to-date documentation

Maintaining up-to-date documentation is fundamental to ensuring that DPIAs accurately reflect the scope of the assessment and effectively fulfil their primary objective: identifying and mitigating risks associated with processing activities. Compliance with the GDPR and broader data protection laws is best achieved through regular, in-depth reviews of existing documentation as a standard practice.

A structured approach requiring each partner to periodically verify the accuracy of the information provided in DPIAs serves as an effective means of embedding the data protection by design principle into the project. Such an approach ensures that compliance obligations are met and that any changes in processing activities are promptly reflected in the assessments.

To encourage continuous updates, particularly in multi-party projects, the use of structured questionnaires has proven to be an effective tool. These questionnaires facilitate a dynamic review process, allowing for a comparative analysis of processing activities at different stages. This approach enables stakeholders to determine whether an updated DPIA assessment is

necessary based on evolving project requirements, technological changes, or regulatory developments.

4.4 HARPOCRATES

The HARPOCRATES project exemplifies best practices in data protection by integrating DPIAs into its technical and organizational design. A key element of the project was the collaborative DPIA threshold assessment, conducted through an interactive workshop-based approach using Mentimeter. This participatory process engaged project partners in contrasting and assessing potential high-risk processing activities against GDPR and EDPB guidelines, fostering a shared understanding of compliance responsibilities across various sectors and disciplines.

To ensure proportionality and necessity, additional safeguards were discussed, including data minimization, pseudonymization, secure data-sharing protocols, and strict retention controls. The project applied privacy-by-design and privacy-by-default principles by incorporating state-of-the-art privacy-enhancing technologies, such as functional encryption, hybrid homomorphic encryption, and federated learning at two levels:

- DPIA-based measures were applied to research data processing. The findings from these assessments were used to mitigate risks associated with processing sensitive data, particularly in contexts involving vulnerable individuals and emerging technologies.
- DPIA-related analysis concerning the tension between GDPR principles and requirements was collaboratively discussed and considered in the design of HARPOCRATES technologies.

The DPIA process also facilitated transparency and accountability, supporting not only legal compliance but also ethical data stewardship throughout the project's demonstrators. By embedding the DPIA as a living, iterative process and engaging stakeholders continuously, HARPOCRATES sets a strong example of how privacy risk assessments can drive innovation while safeguarding fundamental rights.

4.5 TITAN

In the TITAN project, a core best practice is integrating legal and technical perspectives from the earliest project stages, ensuring GDPR requirements and privacy mechanisms are embedded holistically into the technical design. To achieve this, the following strategies have been implemented:

- Continuous Collaboration Between Legal and Technical Teams
  Given the project's focus on advanced technologies (AI, Federated Learning) and sensitive data processing, seamless communication between data protection experts and developers is critical. Legal experts actively participate in technical working groups to identify real-time risks (e.g., in Federated Learning models) and propose mitigation measures aligned with the PEP-PDP framework. This collaboration is reinforced through specialized workshops where GDPR principles—such as "data anonymization"

or "purpose limitation"—are translated into technical requirements for FIWARE-based Data Space connectors.

- Early Data and Flow Mapping
  From the project's initial months, partners were tasked with identifying categories of the sensitive data (e.g., health or biometric data) and documenting data flows across the organizations within the EOSC ecosystem. This exercise enabled the creation of TITAN-specific data-sharing agreements, clarifying roles (data controllers vs. processors) and implementing technical safeguards like automated pseudonymization via Privacy-Preserving Enablers (PPEs).
- Dynamic DPIA Iterations
  The Data Protection Impact Assessment (DPIA) is treated as a living process. Each version includes actionable steps, such as implementing temporary access tokens in the Data Space or updating RBAC policies. Regular reviews ensure the DPIA adapts to technological advancements (e.g., new AI functionalities) or regulatory changes. Technical findings, such as anonymization test results, are directly integrated into risk assessments to maintain alignment with GDPR compliance.
- Integration of Legal-Technical Mechanisms into FIWARE Architecture
  The FIWARE-based Data Space, equipped with PEP-PDP connectors, enables the translation of legal policies into enforceable technical rules. For instance, GDPR consent clauses (Art. 7) are automatically converted into real-time access policies, ensuring compliance is baked into system operations. Audit logs generated natively by the PDP further meet GDPR transparency requirements (Art. 5.2).
- Dual Approach: Technical + Organizational Measures
  Following IDS recommendations, TITAN combines technical and organizational safeguards:
  - Technical measures include homomorphic encryption for sensitive data analysis and anonymization/aggregation modules embedded within PPEs.
  - Organizational measures cover standardized APIs for managing ARCO rights (Art. 15-22 GDPR), RBAC access policies, and consent management protocols (storage and withdrawal). Quarterly data breach simulations ensure readiness for GDPR-mandated notifications (Art. 33-34).
- Proactive Privacy Culture

To foster accountability, developers receive gamified GDPR training using real-world TITAN scenarios (e.g., validating AI models for data minimization). A shared GitHub repository documents technical-legal decisions, enabling cross-functional audits and preventing knowledge silos.

Impact

This methodology ensures TITAN's core components incorporate privacy by design in a verifiable manner. It not only complies with GDPR but also aligns with standards like IDS and the EU Cybersecurity Framework, reinforcing trust in secure data sharing within the EOSC ecosystem.

5. Conclusion

While a DPIA is legally required under the GDPR for specific types of data processing, the projects reviewed go beyond compliance, aiming to foster collective and responsible

innovation —and can benefit from approaching the DPIA as more than a compliance exercise. By integrating DPIAs proactively, these projects demonstrate a commitment to legal accountability, comprehensive reporting to European Commission, responsible use of public funding, and enhance the credibility of their outcomes in the eyes of regulators and the broader scientific community.

The majority of the projects reviewed approached the DPIA not only from a standpoint of legal necessity, but also as a strategic tool to identify potential risks in the processing of sensitive data and to evaluate the effectiveness of safeguards implemented.

DPIAs were applied flexibly—sometimes to assess closely related data processing operations, and at other times to examine broader technologies or systems. Projects often relied on guidance and templates from national Data Protection Authorities.

Key takeaways from these experiences highlight the importance of viewing the DPIA as a continuous and collaborative process (FLUTE, TRUMPET), embedding DPIAs into project lifecycles by involving stakeholders early in the process (WARIFA). Regular training (PAROMA-MED, TITAN) was essential to build a shared understanding of critical concepts, such as anonymization, among diverse project partners. Structured procedures, such as questionnaires (FLUTE, WARIFA, ENCRYPT) to document processing activities and periodic reviews of DPIA content were implemented were implemented to support ongoing alignment and uphold data protection diligence. Partner involvement (TRUMPET, ENCRYPT et al.) proved crucial in mapping data flows and verifying the effectiveness of security measures.

In practice, the projects implemented a range of technical and organizational measures, including pseudonymization, anonymization, privacy-preserving technologies, data minimization strategies, access controls, multi-factor authentication, and activity logging. These efforts underscore that a well-executed DPIA process can serve as both a foundation for compliance and a catalyst for innovation (HARPOCRATES), cultivating a privacy-aware culture within open science.

List of abbreviations:

AI: Artificial Intelligence

CISO: Chief Information Security Officer

CNIL: French Data Protection Authority

CT: Computer Tomography

CTI: Cyber Threat Information

csPCa: Clinically Significant Prostate Cancer

DPA: Data Protection Authority

DPIA: Data Protection Impact Assessment

DPO: Data Protection Officer

DP: Differential Privacy

EOSC: European Open Science Cloud

FE: Functional Encryption

FHE: Fully Homomorphic Encryption

FL: Federated Learning

GDPR: General Data Protection Regulation

HHE: Hybrid Homomorphic Encryption

HTA: Health Technology Assessment

IDS: International Data Spaces

ML: Machine Learning

NLP: Natural Language Processing

NSCLC: Non-Small Cell Lung Cancer

PETs: Privacy-Enhancing Technologies

PIA: Privacy Impact Assessment

PPEs: Privacy-Preserving Enablers

PROMs: Patient-Reported Outcomes

PPT: Privacy-Preserving Technology

RBAC: Role-Based Access Control

RWD: Real-World Data

RWE: Real-World Evidence

SMEs: Small and Medium Enterprises

TEE: Trusted Execution Environment

This work in the project PAROMA-MED is funded by the European Union under Grant Agreement 101070222. The views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission (granting authority). Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union under Horizon Europe Programme (FLUTE GA No. 101095382, TRUMPET GA No. 101070038). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the Health And Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them.

This work is funded in part by the EU-funded ENCRYPT under the Horizon Europe Framework Programme under grant agreement Nr. 101070670

The HARPOCRATES project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101069535. UK participant in Horizon Europe Project HARPOCRATES is supported by UKRI grant number 10048312.

ONCOVALUE is funded by the European Union under Grant Number 101095245. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.

The WARIFA project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017385. The paper reflects only the author's view and that the European Commission is not responsible for any use that may be made of the information it contains.

The TITAN project is funded by the European Union under the GA No 101129822. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

LIST OF EDITORS AND CONTRIBUTORS

AUTHORS	PROJECT	ORGANIZATION
MAIN EDITORIAL TEAM		
POOJA MOHNANI KOSTAS KOUTSOPOULOS ALESSANDRO BASSI	PAROMA-MED	Organization: Eurescom GmbH E-mail: mohnani@eurescom.eu Organization: QualTeK E-mail: k.koutsopoulos@qualtek.eu Organization: Eurescom GmbH E-mail: bassi@eurescom.eu
ELENI MORAITI MAGDALENA KOGUT-CZARKOWSKA	FLUTE and TRUMPET	Organization: TIMELEX E-mail: eleni.moraiti@timelex.eu Organization: TIMELEX E-mail: magdalena.kogut@timelex.eu
CONTRIBUTORS		
MARCO BASSINI	ENCRYPT	Organization: Tilburg University E-mail: m.bassini@tilburguniversity.edu
MARIANO MARTIN ZAMORANO, PHD NENAD GLIGORIC TAMASS KISS	HARPOCRATES	Organization: Trilateral Research E-mail:martin.zamorano@trilateralresearch.com Organization: Zentrix Lab E-mail: nenad.gligoric@zentrixlab.eu Organization: University of Westminster E-mail: T.Kiss@westminster.ac.uk
NEA HELLMAN CHIARA ZOCCHI MIKKO JUVONEN NADIA HONING	ONCOVALUE	Organization: HUS Helsinki University hospital E-mail: nea.hellman@hus.fi Organization: CiaoTech srl E-mail: c.zocchi@ciaotech.com Organization: HUS Helsinki University hospital E-mail: mikko.i.juvonen@hus.fi Organization: Ttopstart E-mail: Nadia.honing@ttopstart.com
CHIARA ZOCCHI MANUELA GUIDUCCI	WARIFA	Organization: CiaoTech srl E-mail: c.zocchi@ciaotech.com Organization: CiaoTech srl E-mail: M.Guiducci@ciaotech.com