



**ANNEX 1**



**Horizon Europe (HORIZON)**

**Description of the action (DoA)**

**Part A**


**Part B**



fdec76f2



Project: 101069535 — HARPOCRATES — HORIZON-CL3-2021-CS-01

 Associated with document Ref. Ares(2022)4525586 - 20/06/2022

**DESCRIPTION OF THE ACTION (PART A)**

**COVER PAGE**

Part A of the Description of the Action (DoA) must be completed directly on the Portal Grant Preparation screens.

<b>PROJECT</b>	
Grant Preparation (General Information screen) — Enter the info.	
<b>Project number:</b>	101069535
<b>Project name:</b>	Federated Data Sharing and Analysis for Social Utility
<b>Project acronym:</b>	HARPOCRATES
<b>Call:</b>	HORIZON-CL3-2021-CS-01
<b>Topic:</b>	HORIZON-CL3-2021-CS-01-04
<b>Type of action:</b>	HORIZON-RIA
<b>Service:</b>	CNECT/H/01
<b>Project starting date:</b>	fixed date: 1 October 2022
<b>Project duration:</b>	36 months

**TABLE OF CONTENTS**

Project summary ..... 3

List of participants ..... 3

List of work packages ..... 5

Staff effort ..... 12

List of deliverables ..... 13

List of milestones (outputs/outcomes) ..... 19

List of critical risks ..... 20

Project reviews ..... 22



Project: 101069535 — HARPOCRATES — HORIZON-CL3-2021-CS-01

 Associated with document Ref. Ares(2022)4525586 - 20/06/2022

## PROJECT SUMMARY

<p><b>Project summary</b></p> <p>Grant Preparation (General Information screen) — Provide an overall description of your project (including context and overall objectives, planned activities and main achievements, and expected results and impacts (on target groups, change procedures, capacities, innovation etc)). This summary should give readers a clear idea of what your project is about.</p> <p>Use the project summary from your proposal.</p> <p>Availability of large volumes of user data combined with tailored statistical analysis present a unique opportunity for organizations across the spectrum to adapt and finetune their services according to individual needs. Having shown remarkable results in analyzing user data, machine learning models attracted global adulation and are applied in a plethora of applications including medical diagnostics, pattern recognition, and threat intelligence. However, such service improvements and personalization based on user data analysis come at the heavy cost of privacy loss. Furthermore, practice showed that systems that use such models incorporate proxies that are often inexact, biased and often unfair. In HARPOCRATES, we focus on setting the foundations of digitally blind evaluation systems that will, by design, eliminate proxies such as geography, gender, race, and others and eventually have a tangible impact on building fairer, democratic and unbiased societies. To do so, we plan to design several practical cryptographic schemes (Functional Encryption and Hybrid Homomorphic Encryption) for analyzing data in a privacy-preserving way. Besides processing statistical data in a privacy-preserving way, we also aim to enable a richer, more balanced and comprehensive approach where data analytics and cryptography go hand in hand with a shift towards increased privacy. In HARPOCRATES we will first show how to effectively combine cryptography with the principles of differential privacy to secure and privatise databases. Next, we will build privacy-preserving machine learning models able to classify encrypted data by performing high accuracy predictions directly on ciphertexts across federated data spaces. Finally, to demonstrate how these solutions respond to users' needs, we will implement two real-world cross-border data sharing scenarios related to health data analysis for sleep medicine and threat intelligence for local authorities.</p>
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

## LIST OF PARTICIPANTS

<b>PARTICIPANTS</b>					
Grant Preparation (Beneficiaries screen) — Enter the info.					
Number	Role	Short name	Legal name	Country	PIC
1	COO	TUNI	TAMPEREEN KORKEAKOULUSAATIO SR	FI	902999288
2	BEN	RISE	RISE RESEARCH INSTITUTES OF SWEDEN AB	SE	999613422
3	BEN	TRI IE	TRILATERAL RESEARCH LIMITED	IE	909169458
4	BEN	ZENTRIX LAB LLC	PRIVREDNO DRUSTVO ZENTRIX LAB DRUSTVO SA OGRANICENOM ODGOVORNOSCU PANCEVO	RS	898982906
5	BEN	CBIT	CANARY BIT AB	SE	891203700
6	BEN	CHARITE	CHARITE - UNIVERSITAETSMEDIZIN BERLIN	DE	999992692
7	BEN	UMG	UNIVERSITAETSMEDIZIN GOETTINGEN - GEORG-AUGUST-UNIVERSITAET GOETTINGEN - STIFTUNG OEFFENTLICHEN RECHTS	DE	999492657
8	BEN	SARGA	SOCIEDAD ARAGONESA DE GESTION AGROAMBIENTAL SL	ES	950464977
9	BEN	VR-ICTdep	REGIONE DEL VENETO	IT	999465691



Project: 101069535 — HARPOCRATES — HORIZON-CL3-2021-CS-01

 Associated with document Ref. Ares(2022)4525586 - 20/06/2022

<b>PARTICIPANTS</b>					
Grant Preparation (Beneficiaries screen) — Enter the info.					
<b>Number</b>	<b>Role</b>	<b>Short name</b>	<b>Legal name</b>	<b>Country</b>	<b>PIC</b>
10	BEN	UEF	ITA-SUOMEN YLIOPISTO	FI	991207984
11	BEN	UP	UNIVERSITE PARIS CITE	FR	897691060
12	BEN	S2 GRUPO	S2 GRUPO DE INNOVACION EN PROCESOS ORGANIZATIVOS SL	ES	986190853
13	AP	999858250	THE UNIVERSITY OF WESTMINSTER LBG	UK	999858250



**LIST OF WORK PACKAGES**

<b>Work packages</b>						
Grant Preparation (Work Packages screen) — Enter the info.						
Work Package No	Work Package name	Lead Beneficiary	Effort (Person-Months)	Start Month	End Month	Deliverable No(s)
WP1	Operations on Encrypted Data and Differential Privacy	1 - TUNI	52.00	1	30	D1.2, D1.1
WP2	Privacy-Preserving Machine and Federated Learning	2 - RISE	72.00	6	30	D2.1
WP3	Reference Architecture & Platform Integration	5 - CBIT	76.00	1	36	D3.1, D3.2
WP4	Use Cases & Demonstrators	13 - 999858250	146.00	6	36	D4.2, D4.1
WP5	Dissemination, Exploitation and Communication	4 - ZENTRIX LAB LLC	45.00	1	36	D5.1, D5.3, D5.2
WP6	GDPR Compliance, Legal and Ethical Impacts of Privacy-Preserving Technologies	3 - TRI IE	49.00	1	36	D6.2, D6.1
WP7	Project Management and Consortium Coordination	1 - TUNI	29.00	1	36	D7.2, D7.3, D7.1



f0dec76f2

Project: 101069535 — HARPOCRATES — HORIZON-CL3-2021-CS-01

 Associated with document Ref. Ares(2022)4525586 - 20/06/2022

### Work package WP1 – Operations on Encrypted Data and Differential Privacy

<b>Work Package Number</b>	WP1	<b>Lead Beneficiary</b>	1. TUNI
<b>Work Package Name</b>	Operations on Encrypted Data and Differential Privacy		
<b>Start Month</b>	1	<b>End Month</b>	30

#### Objectives

In WP1 all the appropriate cryptographic schemes for processing, storing and sharing data in a secure and privacy-preserving way will be developed. In addition to that, the main functionality for performing statistical computations in a privacy-preserving way (i.e. over encrypted data) will be designed. The main objectives of WP1 will be:

- O1.1 Analyze existing HE, HHE and FE schemes and expose current inefficiencies, vulnerabilities and privacy issues;
- O1.2 Design an efficient and secure HHE scheme;
- O1.3 Design several symmetric and asymmetric FE schemes with support for function hiding;
- O1.4 Design several MPC protocols that will be used as building blocks to enhance the applicability and functionality of our HHE and FE schemes;
- O1.5 Add support for differential privacy in our HHE and FE schemes.

#### Description

Task T1.1: Efficient Homomorphic Encryption Schemes. (M1-M20) Leader: RISE.

Contributors: TUNI, UoW

T1.1 will focus on designing an efficient HHE scheme both in terms of homomorphic evaluation time and noise consumption. HARPOCRATES HHE scheme will be designed in such a way that will be combined with the results of T1.3 to allow the computation of the homomorphic function with multiple inputs from different users. This will allow us to minimize the communication cost in comparison to other HE schemes. Finally, the design of HARPOCRATES's HHE scheme will be done in such a way that it can be combined with HARPOCRATES's FE schemes built in T1.2.

Task T1.2: Symmetric and Asymmetric Functional Encryption with Function Hiding.

(M6-M20) Leader: TUNI. Contributors: RISE, UoW

The main aim of T1.2 is to design and implement several FE schemes that will be based on both symmetric and asymmetric cryptography. Symmetric FE schemes will have the advantage of being more efficient while asymmetric FE schemes will be able to support a wider range of functionality. The designed FE schemes will also offer function-hiding support { a property that, to the best of our knowledge, is not yet supported by any existing FE scheme. By enhancing our FE schemes with this property we can further limit the leakage of sensitive information since the description of a function that a user is requesting to run will remain hidden. Finally, the results of this task will be combined with the results from T1.4 in order to show how DP and FE can be used effectively to provide a novel solution for designing encrypted private databases.

Task T1.3: Multiparty Computation Schemes to Support HE and FE. (M12-M30) Leader: TUNI. Contributors: RISE, TRI, CBIT

In this task we will create several multiparty computation protocols that will help us towards our goals on performing privacy-preserving computations on encrypted data. More precisely, we will design MPC protocols that will support functions such as: secure matrix multiplication, secure comparison, secure equality test, secure argmin and secure division. These protocols will be used as building blocks to achieve our goals of performing statistical analysis of encrypted data in a privacy-preserving way as well as to fulfill the objectives of WP2 for privacy-preserving machine learning. Finally, our MPC protocols will be used in combination with both the HHE and FE schemes developed in T1.1 and T1.2 in order to support the multi-client setting.

Task T1.4: Differential Privacy with Cryptography. (M12-M30) Leader: TUNI. Contributors: RISE, CBIT

The main idea of this task derives from the fact that while cryptography ensures the confidentiality of the data in an encrypted database, it does not ensure the privacy of the individual users. Believing that securing and privatizing databases are two equilateral problems, T1.4 will address this problem by designing differential privacy mechanisms guaranteeing the privacy of individuals by ensuring similar outputs of queries irrespective of whether an individual's information is present or absent in the database. Finally, the work of this task will be combined with the work of T1.1, T1.2 and T1.3 to provide an additional layer of security in users' data by adding noise prior to encryption.



Project: 101069535 — HARPOCRATES — HORIZON-CL3-2021-CS-01

 Associated with document Ref. Ares(2022)4525586 - 20/06/2022

### Work package WP2 – Privacy-Preserving Machine and Federated Learning

<b>Work Package Number</b>	WP2	<b>Lead Beneficiary</b>	2. RISE
<b>Work Package Name</b>	Privacy-Preserving Machine and Federated Learning		
<b>Start Month</b>	6	<b>End Month</b>	30

#### Objectives

The aim of WP2 is to create ML models that will allow computations over encrypted data with a high accuracy (i.e. evaluate encrypted data generated from WP1).

O2.1 Create models to perform privacy-preserving feature selection;

O2.2 Create models to perform privacy-preserving image, video and audio classification;

O2.3 Create protocols for privacy-preserving, Byzantine robust federated learning.

#### Description

Task T2.1: Privacy-Preserving Feature Selection (M6-M12) Leader: TUNI. Contributors: TUNI, RISE, UoW  
Existing ML approaches assume that the underlying datasets are pre-processed and clean, with features that have been pre-selected and constructed. However, this is impractical for real-life applications. In this task we are planning to overcome this problem by designing and implementing a novel approach for private feature selection based on the filter method, which will be independent of model training. To demonstrate the feasibility of our approach for practical data science, we will perform extensive experiments with the MPC protocols designed in T1.3 of WP1 where computations will be outsourced to multiple servers, with both semi-honest and malicious adversaries. The results of this task will be documented in D2.1.

Task T2.2: Privacy-Preserving Image, Video and Audio Classification (M6-M30) Leader: TUNI. Contributors: RISE, CBIT, UoW

The main aim of this task is to create models that will be able to classify encrypted images, videos and audio files with a high-accuracy and in an efficient way. To achieve this, we will first study existing approaches for privacy-preserving data classification in the context of machine learning. This analysis will allow us to identify possible inefficiencies and flows. As a next step we will move on with the design and implementation of classification approaches where both HHE and FE from WP1 will be used. HHE will be used for the encryption of original images while FE will be used to perform the actual classification. To achieve our goal, we will extensively use the MPC protocols and the encryption schemes developed in WP1.

Task T2.3: Privacy-Preserving and Byzantine-robust Federated Learning (M9-M30) Leader: CBIT. Contributors: ZEN, TUNI, RISE, TRI

In this task we will design and implement a privacy-preserving and Byzantine-robust Federated Learning scheme with model confidentiality. More precisely, we will design an FL scheme that combines the strengths of existing approaches, by protecting data privacy (through additive homomorphic encryption during parameter aggregation), model confidentiality (using confidential computing for remote model deployment) and Byzantine robustness, based on evaluating and refining aggregation rules such as Krum, Brute and Bulyan. We will evaluate its efficiency and Byzantine robustness using federated data spaces available internally within the project consortium (from the demonstrator partners) and externally within the framework of the Gaia-X project.

### Work package WP3 – Reference Architecture & Platform Integration

<b>Work Package Number</b>	WP3	<b>Lead Beneficiary</b>	5. CBIT
<b>Work Package Name</b>	Reference Architecture & Platform Integration		
<b>Start Month</b>	1	<b>End Month</b>	36

#### Objectives

In this WP we will: (1) Collect and prioritize stakeholder technical and security requirements; (2) Design HARPOCRATES's core architecture; and (3) Define integration endpoints and integrate the components designed and developed in WP1 and WP2. The main objectives of this WP are to:



Project: 101069535 — HARPOCRATES — HORIZON-CL3-2021-CS-01

 Associated with document Ref. Ares(2022)4525586 - 20/06/2022

O3.1 Technical requirements for the development of HARPOCRATES components (HE, FE, PPML, PPFL) and privacy-related requirements to guide the design of the framework;  
 O3.2 Define the reference architecture of HARPOCRATES to support component integration;  
 O3.3 Perform component integration based on the reference architecture, and automate the technical evaluation of the framework using test suites provided with the implemented components.

#### Description

Task T3.1: Technical Requirements and Reference Architecture. (M1-M9) Leader:

CBIT. Contributors: All

Technical and security requirements towards HARPOCRATES will be elicited and defined based on project's use cases and in close collaboration with the demonstrator partners. Moreover, the overall architecture of HARPOCRATES describing its main components, mechanisms, algorithms and models, the interconnection scheme and the specific interfaces for exchanging information among them will be designed and described in detail. This task will produce a detailed requirements report.

Task T3.2: Technical Integration Points and Testing Plan (M9-M18) Leader: CBIT. Contributors: All

Platform integration will complement the reference architecture for the HARPOCRATES components provided as standalone libraries in container format. Furthermore, a testing and technical evaluation plan will describe the test automation infrastructure to support the test suites provided with the components to verify the functional and performance aspects of the components.

Task T3.3: Code Maintenance and Continuous Integration (M13-M36) Leader:

CBIT. Contributors: All

This task will be responsible for the continuous integration of the various security components developed by WP1 and WP2 into final HARPOCRATES Framework. The task will deploy and operate a code maintenance repository with enhanced version control and continuous integration capabilities. Along with the version control system, a software quality evaluation toolset (e.g. Sonar and Moose) will be deployed working in compliance to the operating version control system so as to extract software metrics and detect duplicated code, coding standards, unit tests, code coverage, complex code and potential bugs. To facilitate cloud-based deployment, the developed components will be described with MiCADO deployment descriptors. The results of this task will be documented in a release candidate, followed by a final release.

### Work package WP4 – Use Cases & Demonstrators

<b>Work Package Number</b>	WP4	<b>Lead Beneficiary</b>	13. 999858250
<b>Work Package Name</b>	Use Cases & Demonstrators		
<b>Start Month</b>	6	<b>End Month</b>	36

#### Objectives

The main aim of WP4 is to develop and deploy the two cross-border HARPOCRATES demonstrators in a multi-cloud testbed. More specifically the objectives of WP4 are to:

O4.1 Create a cloud computing testbed including both private and public cloud resources for the deployment and prototyping of the demonstrators;

O4.2 Extend the demonstrators with HARPOCRATES security services in order to showcase how these services improve the security and usability of the demonstrator applications;

O4.3 Test, benchmark and evaluate the developed demonstrators and provide feedback to component and platform developers.

#### Description

Task T4.1: Cloud Computing Testbed (M6-M36) Leader: UoW. Contributors: CBIT, ZEN, RISE

This task will set up and operate a cloud computing testbed where the demonstrator applications will be deployed. The task will assure that a suitable cloud testbed, incorporating both private and public cloud resources is operational and available for hosting the demonstrators. Specific requirements of the demonstrator applications regarding hardware, software, network and other resources will be considered and accommodated.

Task T4.2: Cloud Deployment of Harpocrates Demonstrators (M10-M20) Leader:





UoW. Contributors: All  
 Deploy the demonstrator applications on the cloud computing testbed. Initial deployment will concentrate on the applications representing the state-of-the-art at the start of the project. It will also assure that cutting-edge technologies are utilised for the optimised deployment and operation of the applications. Container technologies and cloud orchestration and monitoring/optimisation solutions will assure optimised deployment and run-time operation of the applications.  
 Task T4.3: Extending Harpocrates Demonstrators with Security Services (M20-M36)  
 Leader: UoW. Contributors: All  
 This task will implement the HARPOCRATES demonstrators that utilize the security services developed by the project. It will also provide valuable feedback to the security component and platform development work packages (WP1-3) for the further refinement of the platform and its building blocks. The outcome of this task will be two high quality demonstrators showcasing the results of the project on relevant real-life healthcare and public sector case-studies.

**Work package WP5 – Dissemination, Exploitation and Communication**

<b>Work Package Number</b>	WP5	<b>Lead Beneficiary</b>	4. ZENTRIX LAB LLC
<b>Work Package Name</b>	Dissemination, Exploitation and Communication		
<b>Start Month</b>	1	<b>End Month</b>	36

**Objectives**

The overall goal of this WP is to maximize the impact of the project through a set of carefully planned communication, dissemination, exploitation and standardization activities, that will facilitate the scientific and technological outcomes of HARPOCRATES to attract its audience, capitalize on developed IPR and standardize most innovative results. The objectives are to:

- O5.1 Dene a communication plan and create assets and materials;
- O5.2 Bootstrap and expand the ecosystem and enable its sustainability beyond the project lifetime;
- O5.3 Disseminate the scientific and technological outcomes through relevant conferences, workshops and journals as well as in standardization bodies and forums;
- O5.4 Support commercial exploitation of project results and secure IPR via development and evaluation of appropriate business models, and elaboration of the exploitation strategy.

**Description**

Task T5.1: Communication & dissemination plan, assets, materials and activities (M1-M36) Leader: ZEN. Contributors: All  
 The task will make use of the European Commission's communication best practices to dene dissemination plans, select appropriate tools to be used by the consortium for both internal and external communication, and carry out communication and dissemination activities. The task will include creating a visual identity for the project (e.g. presentation template, factsheet, logo), creating and maintaining a website and social networking proles, and conducting other dissemination and communication activities (e.g. press releases, leaflets, news items etc.).

Task T5.2: Ecosystem bootstrap and expansion (M1-M36) Leader: ZEN. Contributors: All  
 This task will create the Harpocrates ecosystem by leveraging on existing ecosystems and communities, and thus minimizing the amount of work and resources which are required to contribute strengthening the EU's cybersecurity capacities and sovereignty in digital technologies. The task will establish a focus group concentrating on security in healthcare and public sector settings that advises the project and also act as "evangelists" when spreading information about the project's innovations. Additionally, the task will build links between HARPOCRATES and already existing ecosystems and communities (e.g. the FiWARE association and the OpenStack community) with the aim of supporting the creation of sustainable and secure services for the targeted communities.

Task T5.3: Engaging the scientific community and standardization (M1-M36) Leader: CBIT. Contributors: All  
 Implement a strategy to target and reach the scientific community by identifying publication targets, coordinating scientific publication writing, offering training events, and coordinating project representation at scientific conferences and workshops (e.g. presentations, special sessions, targeted workshops, etc.), where appropriate in collaboration with other EU projects. Additionally, it will also assure that the project contributes to standardisation activities in the areas of cybersecurity and cloud-based services (e.g. IETF (NISEC, RATS, and TEEP WGs), OASIS TOSCA WG, ISO (IEC JTC 1/SC42), and OSF OpenStack Edge Computing Group).

Task T5.4: Innovation management, IPR handling and future exploitation (M1-M36) Leader: ZEN. Contributors: All



Project: 101069535 — HARPOCRATES — HORIZON-CL3-2021-CS-01

 Associated with document Ref. Ares(2022)4525586 - 20/06/2022

This task, going beyond the Consortium Agreement, will detail IPRs and ownership of results and data, to guarantee the rights of all partners in the consortium, in terms of foreground and background knowledge. It will also focus on the sustainability of the results and the generated knowledge after the completion of the project. The plan will include all publicly available assets and will provide directions to partners for the sustainability of these results, a plan for keeping these results alive, and for further extending their impact and reach.

### Work package WP6 – GDPR Compliance, Legal and Ethical Impacts of Privacy-Preserving Technologies

<b>Work Package Number</b>	WP6	<b>Lead Beneficiary</b>	3. TRI IE
<b>Work Package Name</b>	GDPR Compliance, Legal and Ethical Impacts of Privacy-Preserving Technologies		
<b>Start Month</b>	1	<b>End Month</b>	36

#### Objectives

This work package will complement the project by incorporating legal expertise and assessing and ensuring compliance of the technical results with data regulations, GDPR and research ethics.

O6.1 Monitoring of Research Ethics in HARPOCRATES;

O6.2 Systematic review of GDPR compliance automation through privacy-preserving technologies;

O6.3 Devising accessible recommendations for policy-makers and researchers.

#### Description

Task T6.1: Research Ethics (M1-M36) Leader: TRI. Contributors: All

T6.1 covers the adherence to research ethics principles in such areas as data protection and research with human participants in HARPOCRATES. This task will ensure the compliance with EC ethics guidelines and the Research Ethics reports (in D6.1, D6.2) will document the project's conformity to potential ethical issues and the policies and measures that must be followed. It will also provide details for monitoring and compliance activities during the project.

Task T6.2: Automating GDPR Compliance: A Systematic Review (M1-M12) Leader: TRI. Contributors: ZEN

T6.2 provides a systematic analysis of academic and scientific literature on the potential and the shortcomings of privacy-preserving technologies to achieve GDPR compliance. It distinguishes among approaches that offer user-centred data protection, ones that offer automated compliance, and ones that seek to promote learning from big data analysis in a privacy-preserving manner. The systematic review report (D6.1) will highlight best practices in previous research and key gaps to inform the contribution of HARPOCRATES to GDPR compliant data spaces for research and digital services in the EU.

Task T6.3: GDPR Impact Assessment - HARPOCRATES and GDPR Compliance (M12-M32) Leader: TRI. Contributors: All

Informed by T6.2, this task will assess the contribution of HARPOCRATES to GDPR compliance by reference to the ways in which the project will go beyond the state of the art. From the technological innovation perspective, T6.3 will ensure the contribution of technologies, such as HE, PPML and PPFL, to GDPR protection. From a legal analysis perspective, T6.3 will consider the particularities of HARPOCRATES data processing use cases for GDPR protection, offering a comprehensive assessment of how the project's technologies will improve compliance with GDPR principles such as the rights of data subjects not to be subjected to automated decisions. Considering the difficulties with establishing a common threshold for anonymity, the GDPR impact assessment report (in D6.2) will take a broader approach and emphasise the importance of legal safeguards independent from producing legally anonymous data at every step of the processing operations.

Task T6.4: Ethics and Human Rights Impact Assessment (M12-M32) Leader: TRI. Contributors: TRI, ZEN

T6.4 will provide an ethics and human rights impact assessment, articulating how HARPOCRATES balances ethical and human rights safeguards with data use in digital services. While GDPR compliance is very important, there are other fundamental rights and ethical values that need to be considered in this context such as privacy, autonomy, non-discrimination, transparency, accountability, gender equality and environmental sustainability. The Ethics and Human Rights Impact Assessment report

(in D6.2) will detail a comprehensive assessment of technologies within HARPOCRATES that are not only legally compliant, but also designed in a manner that is socially acceptable.

Task T6.5: Policy Recommendations and Synergies for GDPR-compliant data spaces (M32-M36) Leader: TRI. Contributors: TRI, ZEN



Project: 101069535 — HARPOCRATES — HORIZON-CL3-2021-CS-01

 Associated with document Ref. Ares(2022)4525586 - 20/06/2022

T6.5 will build upon the previous tasks of WP6 and will lead to tailor-made recommendations regarding the employment of privacy-preserving technologies to achieve better compliance with GDPR and other fundamental values. The project will also cooperate with projects from topic DATA-01-2021 of Horizon Europe Cluster 4 and other EU-funded and international projects to produce recommendations for the adoption of privacy-preserving technologies to maximise data availability and GDPR compliance. A policy brief (in D6.2) will be produced.

### Work package WP7 – Project Management and Consortium Coordination

<b>Work Package Number</b>	WP7	<b>Lead Beneficiary</b>	1. TUNI
<b>Work Package Name</b>	Project Management and Consortium Coordination		
<b>Start Month</b>	1	<b>End Month</b>	36

#### Objectives

The aim of WP7 is to assure the smooth administrative, financial and technical coordination of the project. To achieve this WP7 has the following objectives:

- O7.1 Creating and running the project management structure and infrastructure including administrative, financial and technical aspects of the project.
- O7.2 Monitoring project progress and resource usage according to Grant Agreement, handling problems or risks identified, coordinating and submitting interim and nal reports to the EC.
- O7.3 Defining, implementing and monitoring the project's Data Management Plan.

#### Description

Task T7.1: Establishing and running the project management structure (M1-M36) Leader: TAU. Contributors: All  
T7.1 will set up and run the project management structure (Project Management Board, Technical Management Board, and Ethical Advisory Committee and deploy its infrastructure (project communication infrastructure, project storage facility and project event management tools.) This task will also organize the project meetings (kick-o and further project meetings), and review meetings.

Task T7.2: Monitoring the project progress (M1-M36) Leader: TAU. Contributors: All  
Monitor project's progress in terms of technical aims, deliverables, milestones, and indicators to be delivered. T7.2 will identify any conflict, delay or risk and will come up with solutions how to handle these issues. This task will be responsible for quality management in the project. T7.2 will coordinate the financial management of the project distributing the budget among beneficiaries according to the Grant Agreement. It will monitor how beneficiaries spend their budget and use their resources. T7.2 will also coordinate compilation of interim and final progress reports collecting inputs from beneficiaries and producing the project's progress reports.

Task T7.3: Defining, implementing and monitoring the project's Data Management Plan (M1-M36) Leader: UoW. Contributors: All  
T7.3 will define the Data Management Plan by creating a detailed outline of the project's policy for data management. This task will create the DMP within the first six months of the project and then implement and refine the plan throughout the project on an ongoing basis.




**STAFF EFFORT**

<b>Staff effort per participant</b>									
Grant Preparation (Work packages - Effort screen) — Enter the info.									
Participant	WP1	WP2	WP3	WP4	WP5	WP6	WP7	Total Person-Months	
1 - TUNI	25.00	20.00	4.00	1.00	1.00	1.00	12.00	64.00	
2 - RISE	15.00	25.00	4.00	1.00	1.00	1.00	1.00	48.00	
3 - TRI IE	2.00	2.00	1.00	1.00	4.00	24.00	1.00	35.00	
4 - ZENTRIX LAB LLC		4.00	6.00	5.00	21.00	7.00	1.00	44.00	
5 - CBIT	5.00	16.00	16.00	4.00	2.00	1.00	1.00	45.00	
6 - CHARITE			5.00	14.00	2.00	2.00	1.00	24.00	
7 - UMG			5.00	14.00	2.00	2.00	1.00	24.00	
8 - SARGA			5.00	20.00	2.00	2.00	1.00	30.00	
9 - VR-ICTdep			5.00	24.00	2.00	2.00	1.00	34.00	
10 - UEF			5.00	14.00	2.00	2.00	1.00	24.00	
11 - UP			5.00	14.00	2.00	2.00	1.00	24.00	
12 - S2 GRUPO			5.00	22.00	2.00	2.00	1.00	32.00	
13 - 999858250	5.00	5.00	10.00	12.00	2.00	1.00	6.00	41.00	
<b>Total Person-Months</b>	<b>52.00</b>	<b>72.00</b>	<b>76.00</b>	<b>146.00</b>	<b>45.00</b>	<b>49.00</b>	<b>29.00</b>	<b>469.00</b>	



## LIST OF DELIVERABLES

<b>Deliverables</b> Grant Preparation (Deliverables screen) — Enter the info. The labels used mean: Public — fully open  automatically posted online) Sensitive — limited under the conditions of the Grant Agreement EU classified — RESTREINT-UE/EU-RESTRICTED, CONFIDENTIEL-UE/EU-CONFIDENTIAL, SECRET-UE/EU-SECRET under Decision <a href="#">2015/444</a>						
Deliverable No	Deliverable Name	Work Package No	Lead Beneficiary	Type	Dissemination Level	Due Date (month)
D1.1	Homomorphic Encryption, Functional Encryption and Function Hiding	WP1	1 - TUNI	R — Document, report	PU - Public	20
D1.2	Multiparty Computation Schemes to Support HE, FE and Differential Privacy.	WP1	2 - RISE	R — Document, report	PU - Public	30
D2.1	Privacy Preserving Feature Selection, Classification and Federated Learning	WP2	2 - RISE	R — Document, report	PU - Public	30
D3.1	Technical Requirements, Architecture and Testing Plan	WP3	2 - RISE	R — Document, report	PU - Public	18
D3.2	Final Release of the HARPOCRATES Framework - Final Release	WP3	5 - CBIT	R — Document, report	PU - Public	36
D4.1	Report on Cloud Computing Testbed	WP4	13 - 999858250	R — Document, report	PU - Public	20
D4.2	Final Demonstrators	WP4	13 - 999858250	R — Document, report	PU - Public	36
D5.1	Plan for dissemination and exploitation including communication activities	WP5	4 - ZENTRIX LAB LLC	R — Document, report	PU - Public	6
D5.2	Initial communication, dissemination, standardisation and exploitation activities	WP5	5 - CBIT	R — Document, report	PU - Public	18
D5.3	Report on communication, dissemination, standardisation and exploitation activities	WP5	4 - ZENTRIX LAB LLC	R — Document, report	PU - Public	36



**Deliverables**

Grant Preparation (Deliverables screen) — Enter the info.

The labels used mean:

Public — fully open (⚠️ automatically posted online)

Sensitive — limited under the conditions of the Grant Agreement

EU classified — RESTREINT-UE/EU-RESTRICTED, CONFIDENTIEL-UE/EU-CONFIDENTIAL, SECRET-UE/EU-SECRET under Decision [2015/444](https://eur-lex.europa.eu/eli/reg/2015/444/2015/444)

Deliverable No	Deliverable Name	Work Package No	Lead Beneficiary	Type	Dissemination Level	Due Date (month)
D6.1	Research ethics report.	WP6	3 - TRI IE	R — Document, report	SEN - Sensitive	18
D6.2	Integrated Impact Assessment Report	WP6	3 - TRI IE	R — Document, report	PU - Public	36
D7.1	Data Management Plan	WP7	13 - 999858250	DMP — Data Management Plan	SEN - Sensitive	6
D7.2	Project interim report	WP7	1 - TUNI	R — Document, report	SEN - Sensitive	18
D7.3	Project final report	WP7	1 - TUNI	R — Document, report	SEN - Sensitive	36



Project: 101069535 — HARPOCRATES — HORIZON-CL3-2021-CS-01

 Associated with document Ref. Ares(2022)4525586 - 20/06/2022

### Deliverable – Homomorphic Encryption, Functional Encryption and Function Hiding

<b>Deliverable Number</b>	D1.1	<b>Lead Beneficiary</b>	1. TUNI
<b>Deliverable Name</b>	Homomorphic Encryption, Functional Encryption and Function Hiding		
<b>Type</b>	R — Document, report	<b>Dissemination Level</b>	PU - Public
<b>Due Date (month)</b>	20	<b>Work Package No</b>	WP1

<b>Description</b>
Documentation that reveals advantages and limitations of existing Homomorphic and Functional Encryption schemes. Furthermore, a detailed analysis of the schemes that will be designed for HARPOCRATES will be presented.

### Deliverable – Multiparty Computation Schemes to Support HE, FE and Differential Privacy.

<b>Deliverable Number</b>	D1.2	<b>Lead Beneficiary</b>	2. RISE
<b>Deliverable Name</b>	Multiparty Computation Schemes to Support HE, FE and Differential Privacy.		
<b>Type</b>	R — Document, report	<b>Dissemination Level</b>	PU - Public
<b>Due Date (month)</b>	30	<b>Work Package No</b>	WP1

<b>Description</b>
Documentation providing a detailed analysis of the MPC protocols that will be designed for HARPOCRATES. In addition to that, detailed protocols that will combine HE, FE and MPC in order to achieve differential privacy will be presented.

### Deliverable – Privacy Preserving Feature Selection, Classification and Federated Learning

<b>Deliverable Number</b>	D2.1	<b>Lead Beneficiary</b>	2. RISE
<b>Deliverable Name</b>	Privacy Preserving Feature Selection, Classification and Federated Learning		
<b>Type</b>	R — Document, report	<b>Dissemination Level</b>	PU - Public
<b>Due Date (month)</b>	30	<b>Work Package No</b>	WP2

<b>Description</b>
Privacy Preserving Feature Selection, Classification and Federated Learning. (M30) Report presenting the implemented privacy-preserving feature selection methods, classification schemes and Byzantine-robust FL schemes.

### Deliverable – Technical Requirements, Architecture and Testing Plan

<b>Deliverable Number</b>	D3.1	<b>Lead Beneficiary</b>	2. RISE
<b>Deliverable Name</b>	Technical Requirements, Architecture and Testing Plan		
<b>Type</b>	R — Document, report	<b>Dissemination Level</b>	PU - Public
<b>Due Date (month)</b>	18	<b>Work Package No</b>	WP3

<b>Description</b>
--------------------



Project: 101069535 — HARPOCRATES — HORIZON-CL3-2021-CS-01

 Associated with document Ref. Ares(2022)4525586 - 20/06/2022

Report describing the HARPOCRATES reference architecture, summarizing the detailed requirements towards the HARPOCRATES Platform, and providing an integration testing and evaluation plan.

### Deliverable – Final Release of the HARPOCRATES Framework - Final Release

<b>Deliverable Number</b>	D3.2	<b>Lead Beneficiary</b>	5. CBIT
<b>Deliverable Name</b>	Final Release of the HARPOCRATES Framework - Final Release		
<b>Type</b>	R — Document, report	<b>Dissemination Level</b>	PU - Public
<b>Due Date (month)</b>	36	<b>Work Package No</b>	WP3

#### Description

A report and demonstrator describing and showcasing the final release of the security components and the overall HARPOCRATES Framework.

### Deliverable – Report on Cloud Computing Testbed

<b>Deliverable Number</b>	D4.1	<b>Lead Beneficiary</b>	13. 999858250
<b>Deliverable Name</b>	Report on Cloud Computing Testbed		
<b>Type</b>	R — Document, report	<b>Dissemination Level</b>	PU - Public
<b>Due Date (month)</b>	20	<b>Work Package No</b>	WP4

#### Description

Describe the architecture, components and operational policies of the cloud computing testbed that will be set up to host the applications and demonstrators. It will also describe the selected healthcare and public sector applications that will be operational on HARPOCRATES cloud testbed.

### Deliverable – Final Demonstrators

<b>Deliverable Number</b>	D4.2	<b>Lead Beneficiary</b>	13. 999858250
<b>Deliverable Name</b>	Final Demonstrators		
<b>Type</b>	R — Document, report	<b>Dissemination Level</b>	PU - Public
<b>Due Date (month)</b>	36	<b>Work Package No</b>	WP4

#### Description

A report and a set of final near production quality demonstrators describing and showcasing the two demonstrator applications, utilizing the final version of the HARPOCRATES Framework. The deliverable will also report on the operational experiences of the cloud testbed and its applications.

### Deliverable – Plan for dissemination and exploitation including communication activities

<b>Deliverable Number</b>	D5.1	<b>Lead Beneficiary</b>	4. ZENTRIX LAB LLC
<b>Deliverable Name</b>	Plan for dissemination and exploitation including communication activities		
<b>Type</b>	R — Document, report	<b>Dissemination Level</b>	PU - Public





Project: 101069535 — HARPOCRATES — HORIZON-CL3-2021-CS-01

 Associated with document Ref. Ares(2022)4525586 - 20/06/2022

<b>Due Date (month)</b>	6	<b>Work Package No</b>	WP5
-------------------------	---	------------------------	-----

<b>Description</b>
Describe the detailed dissemination, communication and exploitation plan of the project (including project website, social media channels and branding), extending the draft plan included in Section 2.

### Deliverable – Initial communication, dissemination, standardisation and exploitation activities

<b>Deliverable Number</b>	D5.2	<b>Lead Beneficiary</b>	5. CBIT
<b>Deliverable Name</b>	Initial communication, dissemination, standardisation and exploitation activities		
<b>Type</b>	R — Document, report	<b>Dissemination Level</b>	PU - Public
<b>Due Date (month)</b>	18	<b>Work Package No</b>	WP5

<b>Description</b>
This will be a report summarizing all activities related to dissemination, communication, ecosystem building, standardization, IPR management and exploitation during the rst half of the project.

### Deliverable – Report on communication, dissemination, standardisation and exploitation activities

<b>Deliverable Number</b>	D5.3	<b>Lead Beneficiary</b>	4. ZENTRIX LAB LLC
<b>Deliverable Name</b>	Report on communication, dissemination, standardisation and exploitation activities		
<b>Type</b>	R — Document, report	<b>Dissemination Level</b>	PU - Public
<b>Due Date (month)</b>	36	<b>Work Package No</b>	WP5

<b>Description</b>
Final report summarizing all activities related to dissemination, communication, ecosystem building, standardization, IPR management and exploitation during the second half of the project.

### Deliverable – Research ethics report.

<b>Deliverable Number</b>	D6.1	<b>Lead Beneficiary</b>	3. TRI IE
<b>Deliverable Name</b>	Research ethics report.		
<b>Type</b>	R — Document, report	<b>Dissemination Level</b>	SEN - Sensitive
<b>Due Date (month)</b>	18	<b>Work Package No</b>	WP6

<b>Description</b>
Report articulating the adherence of HARPOCRATES partners and activities to research ethics requirements, in accordance with Horizon Europe guidance and international best practice.

### Deliverable – Integrated Impact Assessment Report

<b>Deliverable Number</b>	D6.2	<b>Lead Beneficiary</b>	3. TRI IE
---------------------------	------	-------------------------	-----------



Project: 101069535 — HARPOCRATES — HORIZON-CL3-2021-CS-01

 Associated with document Ref. Ares(2022)4525586 - 20/06/2022

<b>Deliverable Name</b>	Integrated Impact Assessment Report		
<b>Type</b>	R — Document, report	<b>Dissemination Level</b>	PU - Public
<b>Due Date (month)</b>	36	<b>Work Package No</b>	WP6

<b>Description</b>
Final report synthesising the insights generated in the GDPR, ethics and human rights impact assessment carried out throughout the HARPOCRATES research.

### Deliverable – Data Management Plan

<b>Deliverable Number</b>	D7.1	<b>Lead Beneficiary</b>	13. 999858250
<b>Deliverable Name</b>	Data Management Plan		
<b>Type</b>	DMP — Data Management Plan	<b>Dissemination Level</b>	SEN - Sensitive
<b>Due Date (month)</b>	6	<b>Work Package No</b>	WP7

<b>Description</b>
It will compile and outline the Data Management Plan of the project.

### Deliverable – Project interim report

<b>Deliverable Number</b>	D7.2	<b>Lead Beneficiary</b>	1. TUNI
<b>Deliverable Name</b>	Project interim report		
<b>Type</b>	R — Document, report	<b>Dissemination Level</b>	SEN - Sensitive
<b>Due Date (month)</b>	18	<b>Work Package No</b>	WP7

<b>Description</b>
Report on all project areas as well as budget and resource usage by beneficiaries.

### Deliverable – Project final report

<b>Deliverable Number</b>	D7.3	<b>Lead Beneficiary</b>	1. TUNI
<b>Deliverable Name</b>	Project final report		
<b>Type</b>	R — Document, report	<b>Dissemination Level</b>	SEN - Sensitive
<b>Due Date (month)</b>	36	<b>Work Package No</b>	WP7

<b>Description</b>
Report on all project areas as well as budget and resource usage by beneficiaries at the end of the project.



**LIST OF MILESTONES**

<b>Milestones</b>						
Grant Preparation (Milestones screen) — Enter the info.						
Milestone No	Milestone Name	Work Package No	Lead Beneficiary	Means of Verification	Due Date (month)	
1	Design and implementation of HARPOCRATES HHE and FE schemes.	WP1	1-TUNI	Completion of Tasks 1.1 and 1.2, release of open source code and delivery of the design of the specific schemes (D1.1).	20	
2	Differential Privacy with Cryptography	WP1	2-RISE	Completion of Task 1.4 delivery of the design of the specific algorithms and release open source code (D1.2).	27	
3	Protocols for privacy-preserving training and features extraction.	WP2	2-RISE	Completion of Task 2.1, delivery of the designed protocols along with the ML models and release of the constructed research artifacts (D2.1).	20	
4	ML models for classification of encrypted data	WP2	1-TUNI	Completion of Task 2.2, delivery of the design of the specific ML models and release of open source code (D2.2).	27	
5	HARPOCRATES Framework released	WP3	5-CBIT	HARPOCRATES Framework released as an opensource set of components.	36	
6	First version of demonstrators deployed	WP4	1-TUNI	First version of the demonstrators operational in cloud testbed (D4.1).	20	
7	Final demonstration	WP4	1-TUNI	Two demonstrator applications are ready using HARPOCRATES Framework (D4.2).	36	
8	Communication, dissemination & exploitation plan	WP5	4-ZENTRIX LAB LLC	Website online and materials prepared. Documented in D5.1.	6	
9	GDPR, ethics and human rights	WP6	3-TRIE	Completion of the initial version of the GDPR, ethics and human rights impact assessment.	18	



**LIST OF CRITICAL RISKS**

<b>Critical risks &amp; risk management strategy</b>			
Grant Preparation (Critical Risks screen) — Enter the info.			
<b>Risk number</b>	<b>Description</b>	<b>Work Package No(s)</b>	<b>Proposed Mitigation Measures</b>
1	The design of a universal HHE scheme may not be possible as its security parameters may not be compatible with ML applications (medium/medium).	WP2, WP1	Restrict the functionalities of HHE by designing a scheme tailored around specific applications and use cases.
2	FE and HE may not be compatible with each other when combined to accelerate the inference phase of a neural network (low/medium).	WP2, WP1	Rely on HHE to decrease the computational costs on the client's side and to maintain at least the same efficiency as approaches that are solely based on HE.
3	A crucial step in designing dynamic private encrypted databases is embedding noise to the data after each update. If not controlled, the noise accumulation can render the database inaccurate (low/high).	WP1	Design of a bootstrap mechanism that periodically reduces the accumulated noise or sets it at the initial level.
4	Building a PPML model where privacy is achieved through training the model on encrypted data seems to be the straightforward solution. However, such an approach can result to extremely high computational costs (high/medium).	WP2	Rely on Split Learning and Federated Learning approaches to train models using plaintext data without sacrificing privacy.
5	New relevant standards and technologies arise making the project technical approach obsolete (medium/high).	WP5, WP3, WP2, WP4, WP1	Literature review as part of each task, and trends within standardization bodies will be monitored throughout the project. Where necessary, technical choices will be re-oriented, considering the latest advances. Strong involvement of research institutions and research-oriented companies, as well as participation in several related fora are expected to ensure rapid identification of such a risk followed by rapid re-orientation where required.
6	Sanitary situation (COVID-19) limits organization of events (medium/low).	WP5	Meetings, conferences and events will be moved to an online format. Increased communication efforts to engage stakeholders getting involved in virtually based knowledge sharing activities.



<b>Critical risks &amp; risk management strategy</b>			
Grant Preparation (Critical Risks screen) — Enter the info.			
<b>Risk number</b>	<b>Description</b>	<b>Work Package No(s)</b>	<b>Proposed Mitigation Measures</b>
7	Low number of audience reached by dissemination and communication activities (medium/high).	WP5	The target audience will be carefully analyzed, and extensive and diverse communication will be tailored to specific end user needs. Liaison with other projects and engagement in common events. Marketing campaigns composed by advertising professionals.
8	Cloud deployment of demonstrators is delayed (medium/medium).	WP4	Automated deployment methods and DevOps best practices will be applied to fasten up development and deployment of the applications.
9	Demonstrators do not illustrate the full capability of the HARPOCRATES Framework and do not reach maturity (medium/medium).	WP4	The demonstrators have been carefully selected to assure that they demonstrate the capabilities of the framework. Requirement collection further synchronises these features. Best practices applied in development to fasten up process.
10	Integration of technical components into HARPOCRATES Framework is delayed (low/high).	WP3, WP2, WP1	Overlap is designed between the outcomes of WP1 and WP2 and the integration activities in WP3. Technical partners of WP1-2 are involved in WP 3 to lower risk.
11	Project objectives are not achieved on time and in required quality due to insufficient resources, external factors (e.g. pandemic), or a partner leaving the consortium (low/high).	WP7	Project progress will be rigorously monitored by coordinator to identify delays and problems early. Coordinating partner (TAU) also has a experience in EU project coordination that helps reducing risk.
12	Partners unwilling to cooperate with ethical and legal requirements (low/medium).	WP6	Compliance with the GDPR and human rights impact assessment framework will be continuously monitored and should any issues arise, the PC will take corrective actions through the project's management structures and procedures.



## **PROJECT REVIEWS**

<b>Project Reviews</b>			
Grant Preparation (Reviews screen) — Enter the info.			
<b>Review No</b>	<b>Timing (month)</b>	<b>Location</b>	<b>Comments</b>
RV1	18	TBC	
RV2	36	TBC	



 Associated with document Ref. Ares(2022)4525586 - 20/06/2022

Version	Date	History of Changes
1	06.04.2022	Part A – WP7: WP7 lead changes from UoW to TUNI, Tasks 7.1 and 7.2 leads changed from UoW to TUNI, UoW effort in WP7 reduced from 13PM to 6PM, TUNI's effort in WP7 increase from 6 PM to 12 PM.
1	06.04.2022	Part A – Deliverables: D7.2 and D7.3 lead passed from UoW to TUNI.
1	06.04.2022	Section 3.2.1 – Change in project coordination has been explained.
1	03.05.2022	Part A – EUR 40,800 Direct Personnel Cost has been transferred from UoW to TUNI to cover the changes in coordination.
1	03.05.2022	Part A: TRI was added to task 1.3 as contributor. ZEN was removed from WP1 as contributor. This change reflects an error in the proposal where the role of the two partners in WP1 was mixed up.
1	06.05.2022	Part A: UoW has been removed as beneficiary and added as Associated partner.
1	06.05.2022	Part A: UoW's effort was reentered to the portal for the relevant work packages and deliverables, after the change in status.
1	06.05.2022	Part A: Responsibility for Milestones 6 and 7 was passed from UoW to TUNI, as the portal does not allow associated partners to take such responsibility.
1	06.05.2022	Part A: UoW researchers were affiliated with the newly added Associated Partner in the portal.



fdec76f2



# Federated Data Sharing and Analysis for Social Utility

## HARPOCRATES\*

### Table of Contents

<b>1 Excellence</b>	<b>3</b>
1.1 Objectives and Ambition.....	3
1.2 Methodology.....	7
<b>2 Impact</b>	<b>19</b>
2.1 Project's Pathway Towards Impact.....	19
2.2 Measures to Maximise Impact.....	22
2.3 Summary.....	29
<b>3 Quality and Efficiency of the Implementation</b>	<b>31</b>
3.1 Work Plan and Resources.....	31
3.2 Capacity of Participants and Consortium as a Whole.....	32
<b>4 Ethics self-assessment</b>	<b>34</b>
4.1 Ethical dimensions of the objectives, methodology and likely impact.....	34
<b>Bibliography</b>	<b>36</b>

---

\*Harpocrates was the god of silence, secrets and confidentiality in the Hellenistic religion.





## 1 Excellence

### 1.1 Objectives and Ambition

Availability of Big Data combined with advancements in Artificial Intelligence (AI) enable broad capabilities for both private and public actors. However, cross-organisation and cross-border data sharing in-line with GDPR is increasingly difficult, as collection of granular, multi-dimensional personal data meets improved capabilities to cross-link data sets. HARPOCRATES leverages novel cryptographic schemes to advance the capabilities of Privacy Preserving Machine Learning (PPML) and Federated Learning (FL), thus enabling decentralised training, validation, and prediction on encrypted data. Such privacy-preserving services and secure computation enable users to both benefit from cloud-based machine intelligence and maintain control over data. HARPOCRATES will enable digitally blind evaluation systems demonstrated in practical application scenarios, helping build fairer, democratic, and unbiased societies.

#### 1.1.1 Project Objectives

We will achieve HARPOCRATES's goals using tools from the domains of cryptography and machine learning (ML). Combining Functional Encryption (FE) and Hybrid Homomorphic Encryption (HHE) will enable efficient processing and classification of encrypted data, beyond the limitations of Homomorphic Encryption (HE). We will further use differential privacy (DP) and Byzantine-tolerant Federated Learning (FL) to provide additional security to data stored and processed in distributed remote locations. Finally, we will develop Privacy-Preserving ML (PPML) models for encrypted data analysis with high accuracy and privacy guarantees. Building these technologies will allow us to develop a cloud-based framework to protect user privacy and limit or even prevent malicious behavior. The framework will provide enhanced data protection and enable users to analyse encrypted data with high accuracy.

**Objective 1: Design efficient function-hiding FE schemes.** Literature on FE considers theoretical aspects of security, ignoring information leakage during execution. We will design FE schemes for statistical functions, where the evaluator is **oblivious of the evaluated function**. See [Table 1.1a](#).

WP	Measurable Indicator	M18	M36
WP1	Formally define a new and stronger <u>threat model</u> for FE.	0	1
WP1	Design function-hiding FE <u>schemes</u> with proofs in new threat model.	1	2
WP3,4	HARPOCRATES FE reference implementation.	v0.5	v1.0
WP1	Peer reviewed publications on HARPOCRATES FE schemes	1	4

Table 1.1a: Measurable indicators of Objective 1

**Objective 2: Combine FE and DP for private encrypted databases.** Combining DP with FE enables privacy-preserving public release of statistics on a dataset. We will design a protocol where encrypted data can be outsourced and queried by many parties, for many functions. See [Table 1.1b](#).

WP	Measurable Indicator	M18	M36
WP2	Structured utility ranking of <u>several</u> DP mechanisms.	5	10
WP2	DP-FE privacy-preserving statistics protocol, <u>low performance penalty</u> .	<50%	<25%
WP3,4	Reference implementation of the HARPOCRATES DP-FE protocol	v0.5	v1.0



Table 1.1b: Measurable indicators of Objective 2

**Objective 3: Design a practical multi-client HHE scheme.** To bypass limitations of HE schemes, we will combine symmetric and asymmetric cryptographic primitives in order to accelerate computationally expensive procedures of HE and design a novel, practical HHE scheme. See [Table 1.1c](#).

WP	Measurable Indicator	M18	M36
WP1	Structured compatibility evaluation of <u>several</u> symmetric-cipher HE schemes.	3	7
WP1	Practical HHE <u>schemes</u> using symmetric and asymmetric primitives.	1	2
WP3,4	HARPOCRATES HHE reference implementation.	v0.5	v1.0
WP1	Peer reviewed publications on HARPOCRATES HHE scheme	1	4

Table 1.1c: Measurable indicators of Objective 3

**Objective 4: Build a PPML framework by combining FE and HHE.** State-of-the-art PPML approaches inefficiently use HE functionality to classify encrypted data with good accuracy. We will combine HHE and FE to achieve faster classification with accuracy on par with state-of-the-art approaches. This will enable privacy-preserving classification of encrypted audio and video files. See [Table 1.1d](#).

WP	Measurable Indicator	M18	M36
WP2	Novel approach for private feature selection.	1	4
WP2	High accuracy PPML model for encrypted file classification.	90%	97%
WP2	Hybrid FE-HHE protocol for classification of encrypted data.	0	1
WP2	Hybrid FE-HHE encrypted data classification protocol, <u>more efficient</u> than published state of the art.	30%	50%
WP2	Peer reviewed publications on privacy-preserving data classification	1	3

Table 1.1d: Measurable indicators of Objective 4

**Objective 5: Byzantine-robust FL scheme with data privacy guarantees.** State-of-the-art FL selectively protects the privacy of training data and model updates, ignoring Byzantine fault tolerance and confidentiality of aggregated models. We will identify approaches to provide formal data security and privacy for FL systems, ensuring robustness to Byzantine faults. See [Table 1.1e](#).

WP	Measurable Indicator	M18	M36
WP2	Byzantine-tolerant FL <u>scheme</u> with global data privacy.	0	1
WP2	Reference implementation for model confidentiality in FL architectures.	v0.5	v1.0

Table 1.1e: Measurable indicators of Objective 5

**Objective 6: Real-world case studies and contribution to Open Science and Reproducible Research.** To ensure applicability of the technical solutions, we will develop two realistic cross-border demonstrator applications in the areas of (i) sleep medicine and (ii) threat intelligence exchange for local authorities. To enable reproducing, extending and enhancing the schemes and the demonstrators developed in HARPOCRATES, we will make our core research results and the developed source code publicly available and will build an active community around these. See [Table 1.1f](#).

WP	Measurable Indicator	M18	M36
WP4	Improved and secure data analytics/ML algorithms on sleep recordings.	1	4
WP4	Privacy-preserving threat intelligence exchange schemes for local public authorities.	1	4



WP5	HARPOCRATES open-source community with registered members and contributors.	50	150
WP4	Peer reviewed publications on HARPOCRATES demonstrators.	1	4

Table 1.1f: Measurable indicators of Objective 6

**Objective 7: Contribute to Scalable Automated GDPR Compliance.** We will identify key challenges preventing technology transfer of privacy-preserving technologies. We will use this insight to facilitate technology transfer of results obtained in HARPOCRATES to help automated GDPR compliance as well as cross-border, privacy-preserving and ethical data processing. See [Table 1.1g](#).

WP	Measurable Indicator	M18	M36
WP6	Embedding of GDPR principles in the design of HARPOCRATES technologies to contribute towards automating GDPR compliance for cross-border data processing	Report	Solution
WP6	Recommendations for the employment of HARPOCRATES solutions by data controllers based on GDPR, ethics and human rights impact assessment	Preliminary version	Final version

Table 1.1g: Measurable indicators of Objective 7

### 1.1.2 Ambition

Pursuing objectives defined in [subsection 1.1.1](#), HARPOCRATES will advance the state of the art as follows:

#### Domain#1: Functional Encryption

**Current Status:** Functional Encryption was introduced in [9] and later formally defined as a generalisation of public-key encryption in [10], followed by extensive research [5–7, 13, 26–28]. Despite promising published work, FE schemes do not support specific functions. To the best of our knowledge, currently supported functionality is limited to inner products [1–3] and quadratic polynomials [41].

**Progress Beyond State-of-the-Art:** We plan to (1) design novel FE schemes in symmetric and asymmetric key settings, supporting a rich set of statistical functions; (2) define a novel threat model applicable to any FE scheme, considering information leakage during the FE scheme runtime; (3) design approaches for generic conversion of message-private multi-input FE schemes to function-private schemes; (4) prove the efficiency of proposed schemes through theoretical and experimental evaluations.

#### Domain#2: Hybrid Homomorphic Encryption

**Current Status:** Homomorphic Encryption was first introduced in [40], allowing limited operations on encrypted data and a breakthrough in [23] allowed any computation on encrypted data (fully homomorphic encryption). Follow-up work [11, 12, 17, 22] failed to meaningfully reduce the performance penalty. Hybrid Homomorphic Encryption (HHE) overcomes the inefficiency of HE schemes [37] by encrypting data with symmetric encryption and outsourcing the symmetric ciphertexts.

**Progress Beyond State-of-the-Art:** We will design new symmetric ciphers optimised for large integer HHE use cases, such as ML. Our symmetric cipher will encrypt plaintexts in  $Z_p$ , for a large prime  $p$ , as ML applications require  $p \gg 2$ . Our cipher will extend state-of-the-art HE schemes (BFV [11, 22], BGV [12]). We will demonstrate theoretically and experimentally the cipher's efficiency.

#### Domain#3: Privacy-Preserving Machine Learning (PPML)

**Current Status:** PPML models were first implemented using Multiparty Computation (MPC), where parties jointly compute a function while keeping inputs private. Lately, we have seen some further developments with two problems being the most important ones: (1) preserving the privacy of the



dataset even when we offload the training or the testing phase to a third party [42, 47, 48] and (2) classifying encrypted data with high accuracy – as if it was unencrypted [25, 33, 46]. The main goal is to reduce leakage of data patterns across client communications while maintaining model accuracy.

**Progress Beyond State-of-the-Art:** HARPOCRATES PPML will be divided into two categories: (1) training and evaluating encrypted data and (2) using only raw data. For the first case, we will create models able to classify encrypted data with high accuracy. Our approach is expected to outperform other similar works in the area since the encryption of the data will be based on our FE and HHE schemes. For the second case, we will build novel models by combining the strengths of Federated Learning (FL) and Split Learning (SL) – the speed of FL and privacy guarantees of SL, which also allows training datasets in a privacy-preserving way with low computing resources.

#### Domain#4: Differential Privacy

**Current Status:** Differential privacy was formalised in [20] for individual privacy: by adding well-calibrated noise to data, presence or absence of an individual's information is *irrelevant* to the output of a query. Modern applications require frequent updates to data, and later models considered scenarios such as real-time traffic analysis, social trends observations and disease outbreaks discovery [21].

**Progress Beyond State-of-the-Art:** We will combine DP with our FE and HHE schemes to build private encrypted databases, a problem that has been widely overlooked. Such a database will protect the confidentiality and privacy of released analytics data against malicious adversaries.

#### 1.1.3 Positioning and Technology Readiness

HARPOCRATES will aim for medium to high Technology Readiness Levels (TRL). The project will extend and implement cybersecurity services built on DP, FE, HE and MPC to preserve privacy in FL. Cryptographic services, together with the HARPOCRATES framework will reach TRL5 and TRL6 by the end of the project and will be demonstrated in industry-relevant environments via the pilot use cases. The demonstrator applications will extend existing implementations and prototypes with HARPOCRATES services and will reach TRL6 and TRL7. We deem target TRLs as realistic and achievable considering that HARPOCRATES will extend previous projects and will aggregate expertise in Distributed Computing, Security, Applied Cryptography, Healthcare and Public Services. Table 1.1h lists the HARPOCRATES security services, framework and demonstrators, along with their TRL levels and respective IPR owner.

Technology	Description	Now	Goal
<b>Symmetric and Asymmetric Functional Encryption</b>	TUNI designed several symmetric and asymmetric FE schemes for privacy-preserving analysis of encrypted data. The schemes work for <i>positive</i> integers and uses MPC to eliminate the need for a trusted party TA; we next plan to support more functions (e.g. quadratic)	TRL2	TRL6
<b>Hybrid Homomorphic Encryption (HHE)</b>	TUNI has designed a plethora of provable secure protocols based on traditional HE. Additionally, it has conducted a study analysing symmetric schemes that can be used for building HHE schemes. Next, we will design new symmetric ciphers optimised for large integer HHE use cases.	TRL2	TRL5
<b>Statistically Analysing Encrypted Data</b>	TUNI and RISE developed a protocol for statistical analysis on encrypted data using FE. We will add support for statistical functions while at the same time MPC will be used to remove the need for a TA.	TRL2	TRL6
<b>Classification of Encrypted Data</b>	TUNI implemented a system for privacy-preserving, high-accuracy identification of content of homomorphically encrypted images. We will next (1) increase efficiency by combining FE and HHE and removing HE, (2) improve model privacy using split learning and federated learning.	TRL2	TRL6
<b>Multi-Party Computation</b>	TUNI designed an MPC protocol that allows the removal of a Trusted Entity in any protocol. This will be further enhanced and will be used in HARPOCRATES's ML models as well as the designed encryption schemes.	TRL2	TRL6



<b>PPFL</b>	RISE built protocols using federated learning. We next plan to improve privacy by protecting the model during the training and the testing phases.	TRL2	TRL5
<b>HARPOCRATES framework</b>	The modular HARPOCRATES Framework aims to incorporate the various security services of the project into an easily deployable and applicable toolkit. CBIT and UoW will integrate the various services into a coherent framework, utilising tools (e.g. MICADO Cloud Orchestrator ) and methodologies successfully applied in earlier EU projects (e.g. ASCLEPIOS).	TRL1	TRL6
<b>Sleep Disorder Demonstrator</b>	CRT, UMG, UEF and VIFASOM will integrate HARPOCRATES services supporting secure collaborative ML into the currently existing Sleep Medicine Analytics Platform for cross-border collaboration. This will help significantly increase the security and usability of the platform.	TRL5	TRL7
<b>Threat Intelligence Demonstrator</b>	SARGA, VENETO and S2 will develop a platform for sharing and analysing threat intelligence information between two European regions.	TRL2	TRL6

Table 1.1h: Baseline and Expected TRLs

### 1.2 Methodology

HARPOCRATES intends to change how organisations and online services store, retrieve, share and process sensitive data. HARPOCRATES will design several new cryptographic techniques and will combine them with ML to deploy accurate and efficient services capable to analyse encrypted data in a fully privacy-preserving manner. Figure 1 presents a coarse-grained overview of HARPOCRATES’s main functionalities, the key players, and the steps involved in the overall process. As indicated in the figure, HARPOCRATES’s goal stretches beyond designing a framework for processing statistical data and releasing statistics in a privacy-preserving way. It will **also provide a richer and more sustainable** approach to data analytics and cryptography, **correcting the power balance between data controllers and data processors**, and one that supports and accelerates the shift towards stronger privacy.

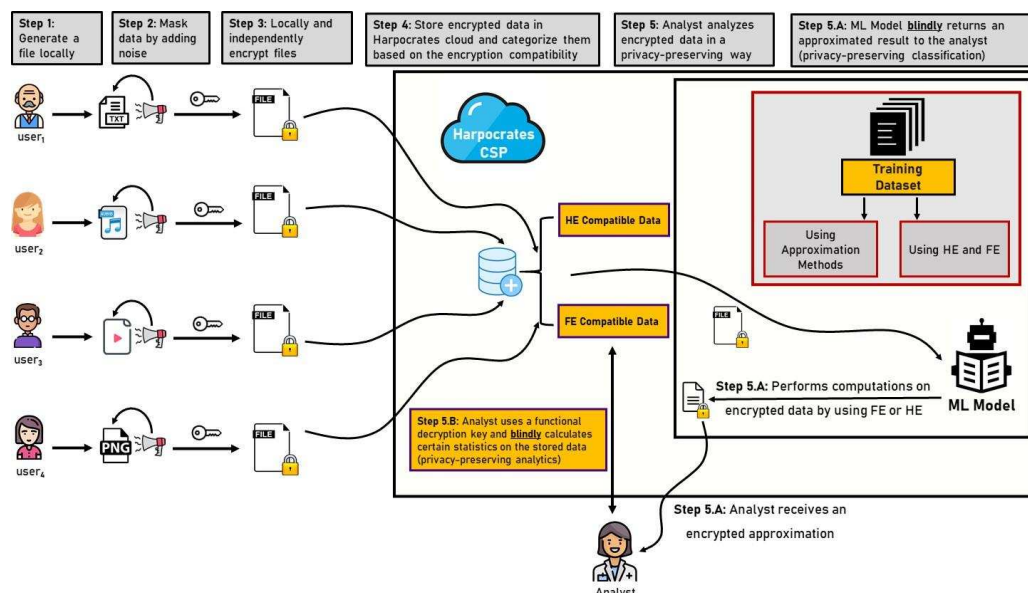


Figure 1: HARPOCRATES High-Level Overview



### 1.2.1 Research Challenges and Enabling Technologies

Analytics and data processing have been gainfully applied in fields as diverse as healthcare (e.g. medical diagnosis), intelligence analysis, finance, safety, military services and many more. However, privacy-preserving analytics has lately gained momentum and broad public attention, an existing trend hastened by the rapid proliferation of COVID-19 contact tracing apps in many countries. As a result, organisations are moving towards implementing services that respect users' privacy.

While there are many works claiming to analyse data in a privacy-preserving way, most of them rely on *non-private* methods such as anonymisation of data<sup>1</sup>. Moreover, as ML revolutionizes various societal domains it also raises serious privacy concerns. For example, data collected for a specific purpose and that is seemingly innocuous to a human can produce far-reaching insights and predictions that only a machine could infer [24]. These observations along with related studies [51] confirmed that citizens care about data privacy and highlighted the need to create more robust and practical privacy-preserving models. With this call to action, researchers started exploring ways to combine breakthroughs in cryptography and machine learning to create the basis for scalable and practical privacy-preserving machine learning (PPML), which has the potential to limit data breaches and privacy violations.

In HARPOCRATES, we will **combine cryptographic methods in PPML to improve efficiency and provide strong guarantees about data privacy**. Furthermore, we will build the infrastructure to enable secure and privacy-preserving data access. Through these efforts we aim to increase the net positive effect of machine learning for society in the coming decades. We broadly classify HARPOCRATES's research into three categories: (1) Encryption schemes that will allow privacy-preserving operations on encrypted data, (2) Improved methods for PPML, and (3) improved applications of PPML. HARPOCRATES focuses on six research topics (R1-R6) that raise a plethora of challenges (C1-C6):

**R1. Functional Encryption [O1-Table 1.1a]:** Functional Encryption (FE) is an emerging cryptographic technique that allows selective computations over encrypted data. FE schemes provide a key generation algorithm that outputs decryption keys with remarkable capabilities. More precisely, each decryption key  $sk_f$  is associated with a function  $f$ . In contrast to traditional cryptographic techniques, using  $sk_f$  on a ciphertext  $\text{Enc}(X)$  does *not* recover  $X$  but a function  $f(X)$  –thus keeping the actual value  $X$  private. While the first constructions of FE allowed the computation of a function over a *single* ciphertext, more recent works [26] introduced the more general notion of multi-input FE (MIFE). In a MIFE scheme, given ciphertexts  $\text{Enc}(x_1), \dots, \text{Enc}(x_n)$ , a user can use  $sk_f$  to recover  $f(x_1, \dots, x_n)$ . The function  $f$  can allow only highly processed forms of data to be learned by the functional key holder. Unfortunately, while MIFE seems to be a perfect fit for many real-life applications – especially cloud-based ones where multiple users store large volumes of data in remote and possibly corrupted entities – most of the works in the field revolve around constructing *generic* schemes that do not support specific functions. Hence, while the concept of FE has the potential to unleash new, creative, useful and emerging applications, from a practical perspective, it still holds a *largely unfulfilled* promise. Having identified the importance of FE and believing that it is a family of modern encryption schemes that can push into an uncharted technological terrain, in HARPOCRATES we will smooth out the identified asymmetries between theory and practice by designing practical symmetric and asymmetric FE schemes. Furthermore, we will show how these schemes can be used as standalone components to perform privacy-preserving computations on encrypted data while at the same time we will demonstrate how these schemes can be combined with ML to design more efficient PPML models.

**C1.** Building secure and practical FE schemes is a complex and difficult task. The first challenge that we will consider in HARPOCRATES, is how to **build efficient symmetric and asymmetric FE schemes to support a wide range of statistical functions**. Next we will improve the security of our schemes by minimising the leakage associated with both the user's query and the actual computation of the functions. To do so, we must ensure that our FE schemes will be **function-hiding** in the sense that the CSP will output the correct result, **without learning anything**

<sup>1</sup>anonymised data can be de-anonymised, also known as re-identification attacks [29].





**about the computed function.** Another important challenge, currently not addressed in the literature, is designing a mechanism allowing users to explicitly specify the input of a function. In the standard FE model the function is applied to all of the users' data. However, this may be extremely problematic in many cases, when the function is not defined over some of the data.

- R2. Differential Privacy [O2–Table 1.1b]:** Differential Privacy (DP) allows sharing information about a dataset, while simultaneously withholding information about individuals. A curator (data owner) creates the database and then periodically releases statistics upon receiving request from an analyst. To ensure the individuals' privacy, the curator filters the statistics through a privacy mechanism and replies to the analyst with a noisy result. The results must be presented in a form allowing the analyst to deduce accurate enough results about the dataset, without breaching individuals' privacy. While the problem of privatizing datasets has been thoroughly studied, further securing the datasets through the use of cryptography has not yet drawn much attention. However, this is an issue of paramount importance when the database is outsourced to a possibly malicious CSP. To the best of our knowledge, the only work that considered this scenario is the one presented in [4], where authors rely on homomorphic encryption [38] and structured encryption [31] to design a scheme for private histogram queries. In HARPOCRATES, we will approach a similar problem by using FE as the starting point and then by including DP in the developed PPML models.
- C2.** While DP is a powerful tool for preserving the privacy of individuals, it currently suffers from important inefficiencies that will be addressed within the framework of HARPOCRATES. The first challenge that we will address is the design of a private encrypted database assuming a stronger threat model than the one presented in current literature. More precisely, in recent state-of-the-art approaches, the role of embedding well-calibrated noise to the ciphertexts, is given to the CSP. As a result, **the security of such approaches is only satisfied under the assumption of an honest CSP.** In HARPOCRATES, we will design schemes considering a malicious CSP. As a next step we will focus on the problem of **minimising of the total accumulated noise after a sequence of updates in the database.**
- R3. Homomorphic Encryption [O3–Table 1.1c]:** Homomorphic Encryption (HE) is an encryption technique that allows users to perform computations on encrypted data without corrupting their features or format. Given two ciphertexts  $c$  and  $c'$ , a user can compute  $f(c, c')$  where  $f$  is a function associated either with an addition or multiplication operation. Many privacy-preserving applications which employ HE use the following design principle. First, the data holder encrypts their dataset using an HE scheme and sends the ciphertexts to a server. The server then performs the computations on the ciphertexts and produces an encrypted result. Only the data holder knows the secret decryption key, so the server has to send the encrypted result to the data holder who can then decrypt it to get the final result of the computation. While this approach protects both the privacy of the input data and the secrecy of the applied computations, it comes with important drawbacks: a drastic performance penalty and ciphertext expansion. This makes ciphertexts in HE schemes several orders of magnitude larger than the corresponding plaintexts. This expansion increases the data volume which has to be transferred from the data holder to the server. This expansion can have a considerable impact on the overall performance of the application, especially on resource-constrained, embedded devices. In HARPOCRATES, we will implement a Hybrid Homomorphic Encryption (HHE) scheme aiming to create practical applications that will utilise the power of HE. HHE was first mentioned in [37]. The main idea behind HHE is the following: Instead of encrypting the data with an HE scheme, encrypt the data with a symmetric cipher (expansion factor of 1) and send the symmetric ciphertexts to the server. The server then first homomorphically performs the symmetric decryption circuit to transform the symmetric ciphertext into a homomorphic ciphertext and then proceeds with performing the actual computations. By building such a scheme, we aim to overcome the main limitations of traditional HE schemes and will show how current and future services can rely on HHE to protect users' privacy. Apart from that, in HARPOCRATES, we plan to use HHE along with numerous machine learning



algorithms in an attempt to classify encrypted data with high accuracy (close to unencrypted). HHE will be used for training and classifying encrypted data using cutting edge ML algorithms such as deep neural networks, support-vector machines, XGBoost and nearest neighbour methods. Finally, in HARPOCRATES, HHE will be also used in combination with FE in order to provide a more efficient solution when classifying encrypted data through the aforementioned algorithms.

- C3.** Homomorphic Encryption is a powerful encryption technique dubbed as “*the Holy Grail of Cryptography*”. However, it is unfortunately inefficient [37]. To bypass this problem, we aim at designing an HHE scheme, by combining a symmetric-key encryption scheme with HE. However, this is not a trivial task as traditional symmetric schemes are not compatible with HE, mainly due to their large multiplicative depth. To this end, the first step of our research will revolve around comparing the compatibility of different symmetric schemes with HE. Consequently, we plan to design a symmetric scheme tailored around the needs of HE, with a strong focus on large integer arithmetic. This will allow us to address a large number of HE use cases and show how our HHE scheme can actually be used to compute **functions of practical interest on encrypted data**. In addition to that, we aim at modifying our HHE scheme in such a way that it will also be compatible with FE. This will be a remarkable result as it will lead to the development of a universal cryptographic scheme, allowing users to combine it either with FE or HE.
- R4. Secure Multiparty Computation [O4–Table 1.1d, O5–Table 1.1e]:** Secure Multiparty Computation (MPC), allows two or more parties to jointly compute a specified output from their private information in a distributed fashion, without mutually revealing their private information. In HARPOCRATES, we will utilise the power of MPC in two ways: (1) we will use MPC to transform our FE schemes from single-client to multi-client. This is an important property since our scheme will allow several users to provide input data encrypted with different keys. (2) MPC will support our PPML models by creating protocols for privacy-preserving feature selection. Feature selection is the process of selecting a subset of relevant features for model training [15]. Using a well chosen subset of features can lead to more accurate models, as well as efficiency gains during model training. A commonly used technique for feature selection is the filter method where features are ranked according to a score indicative of their predictive ability, and subsequently the highest ranked features are retained. Despite its shortcomings (considering each feature in isolation and ignoring feature dependencies) the filter method is popular in practical data science because it is computationally efficient and agnostic of ML model architecture. The designed protocols will enable privacy-preserving feature ranking without needing to pre-process data in a dataset – an important part of the ML model development pipeline largely overlooked in the PPML literature.
- C4.** While MPC is a well-studied field, the problem of combining it with cutting-edge technologies has been widely overlooked in the literature. In HARPOCRATES, we aim to deploy MPC protocols in combination with federated and split learning techniques, in an attempt to design ML applications with enhanced security and stronger privacy guarantees. Apart from that, in the field of FE, and in contrast with current state-of-the-art literature where an unrealistic fully trusted party generates and distributes functional decryption keys, MPC techniques can offer users the ability to generate those keys themselves and thus obviate the need of a fully trusted third party. Hence, in HARPOCRATES we will examine how to utilise MPC to eliminate the need for any trusted authority.
- R5. Privacy-Preserving ML [O4–Table 1.1d]:** In HARPOCRATES, we will illustrate the use of deep neural network algorithms over encrypted data and show how this can protect users’ data and privacy. We will use a combination of FE and HE to perform computations on encrypted data. HE can perform arithmetic operations (addition and multiplication) over encrypted data without decrypting it, allowing any function that uses these arithmetic operations to be homomorphically evaluated. Beyond earlier works utilising HE for classification of encrypted data [14, 18, 25, 30], in HARPOCRATES we will design a **novel hybrid approach** that will utilise HHE for the encryption of the initial data but then use FE for the actual classification. This, will result in a **significantly more efficient algorithms without sacrificing accuracy**. Finally, our PPML models will be used to classify encrypted data of different types (text, images, videos and sound). Finally, the





- applicability of our models will be evaluated in a wide range of fields through our demonstrators.
- C5.** The main challenge with privacy-preserving ML, is that **HE and FE schemes do not currently provide support for non-linear functions**. To this end, in HARPOCRATES we will focus on **finding the best possible polynomial approximations for the activation functions used in ML**. Apart from that, we seek to explore the possibility of designing privacy-preserving models for the classification of encrypted files (image, audio and video) – a problem we believe will make a real difference in providing guarantees to end users about their privacy.
- R6. Privacy-Preserving Federated Learning [O5–Table 1.1e]:** Federated learning and decentralised deep learning allow to process large information volumes faster [49] and generate accurate models [43], using primarily *Model parallelism* and *Data parallelism* [44]. It is tempting to believe that sharing gradients, model updates, or meta-level information (such as outputs of layers in neural networks) in place of raw data protects privacy. However, earlier work demonstrated that gradients exchanged during distributed training leak information about training data [52], model activations [50], and updates [35] - though exchanging model weights instead of gradients offers better privacy [39]. Approaches to protect privacy in decentralised learning includes differentially private mechanisms at training, multi-party secure aggregation to hide individual contributions to the aggregator, additively homomorphic encryption allowing the aggregator to sum updates [45] and combinations thereof. A crucial metric for decentralised machine learning architectures is their *robustness* to failures causing digression from expected behaviour, including *Byzantine failures* [32] where a subset of workers behave arbitrarily and send incorrect gradients to peer workers, thus preventing model convergence when the aggregation rule used by the workers cannot tolerate even a single byzantine gradient. Several *Gradient Aggregation Rules* (GAR's) address this problem by providing  $(\alpha, f)$ -Byzantine Robustness [8]: Brute [36], Krum [8], DRACO [16] and Bulyan [36]. In HARPOCRATES we will combine the cryptographic approaches to preserve privacy in federated learning (mentioned above) with algorithmic approaches for Byzantine-resistance, producing a scalable federated ML architecture that both ensures privacy and is robust to Byzantine inputs.
- C6.** Despite a promising outlook for analysing data in decentralised settings and federated data spaces, Federated Learning (FL) presents **important challenges in three dimensions**: data privacy, model confidentiality, and robustness to Byzantine attacks. In HARPOCRATES we will design an FL scheme combining existing approaches to protect data privacy (through multi-party secure aggregation), model confidentiality (with confidential computing) and Byzantine robustness.

## 1.2.2 Architecture

HARPOCRATES will produce an open extensible architecture, with robust, reusable and self-contained components. The HARPOCRATES architecture consists of six discrete layers described below (Figure 2): Cloud Infrastructure, Crypto, ML, DP, GDPR Compliance, Legal & Ethics, and Demonstrators.

**L1. Cloud Infrastructure Layer:** The Cloud Infrastructure Layer will assure that the research outcomes of HARPOCRATES can be hosted, prototyped and executed in a realistic multi-cloud infrastructure. It will also provide the necessary testbed for the implementation of the demonstrators. The cloud computing testbed will be set up using state of the art technologies for cloud application orchestration to ensure the optimised deployment and execution of the HARPOCRATES components and the demonstrator applications. The targeted multi-cloud testbed will incorporate both private cloud resources based on the OpenStack cloud computing infrastructure of UoW and RISE's ICE Data Centre, and public cloud resources purchased on demand from leading cloud providers such as Amazon AWS, Microsoft Azure or Google Cloud Platform. Using such large variety of resources in the testbed will assure the generic nature and wide applicability of the developed solutions. To provide easy deployment, run-time management and flexible portability between the various cloud platforms, HARPOCRATES will utilise the MICADO [19] application-level cloud orchestration solution that was developed in previous European projects and that has already been successfully applied by some HARPOCRATES partners in the H2020 ASCLEPIOS project. Providing



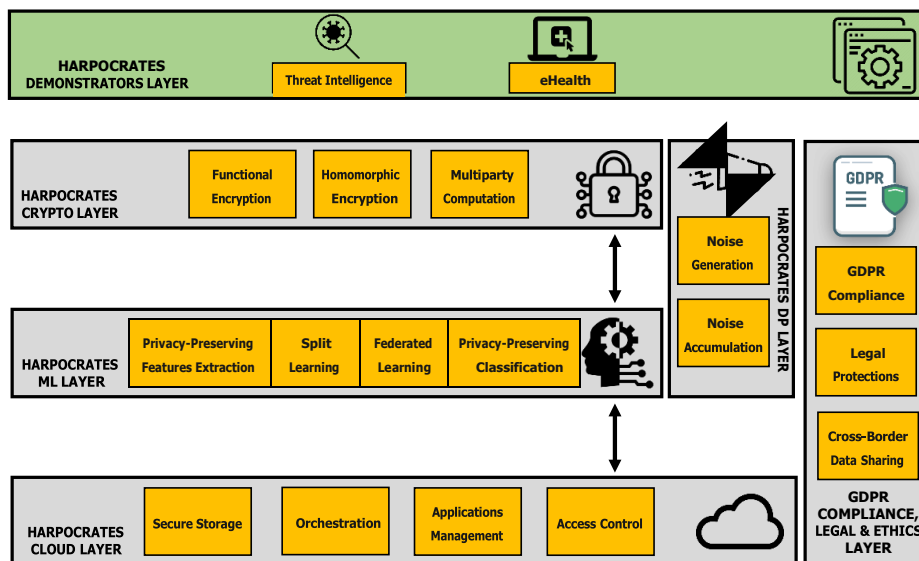


Figure 2: HARPOCRATES Core Layers

MiCADO compatible application descriptors in the standardized TOSCA [34] format will assure that all security and demonstrator application components (implemented as microservices) can be automatically deployed and migrated on-demand between various cloud infrastructures.

**The novelty of this layer** is provided by the flexible and dynamic nature of the microservices architecture, and the way the various application components (microservices) are described. All HARPOCRATES components will be provided as MiCADO application descriptors, made publicly available in open code repositories such as GitHub. Such descriptors can then be composed, either manually or automatically into complex applications, incorporating both the application components of the demonstrators and the additional security services provided by HARPOCRATES. This modular approach will assure easy reproducibility and applicability of the project’s results.

- L2. Crypto Layer:** The crypto layer will comprise a collection of cryptographic algorithms and will be a key component for both the security, privacy and the main functionality of HARPOCRATES. This layer will provide a complete cryptographic toolkit that will be used to protect stored data and offer a plethora of mechanisms for analysing encrypted data in a privacy-preserving way (i.e. without leaking any valuable information about the actual content of the analysed data). The core of the Crypto layer will be an implementation of a Functional Encryption library and a Hybrid Homomorphic Encryption scheme. Regarding FE, in HARPOCRATES we will design both symmetric and asymmetric schemes that, unlike existing work, will support a wide range of statistical functions. In addition to that, HARPOCRATES FE schemes will be function-hiding. That is, to make sure that an entity (e.g. the CSP) who is performing a functional decryption, will output the correct result, without learning anything about the computed function. This is a very important security property that has been overlooked in the current state-of-the-art. In HARPOCRATES we will design an HHE scheme to support practical applications that will utilise the power of HE in an efficient way. HHE will allow users to encrypt their data with a symmetric cipher and send the generated ciphertexts to the server. The server will next transform the symmetric ciphertexts into homomorphic ciphertexts. With this scheme, we will manage to bypass several drawbacks of HE that have prevented this promising technology from being widely adopted. Finally, the Crypto layer will also contain a list of MPC protocols that will be used as a complement to the main HARPOCRATES encryption schemes (i.e. FE and HHE). More precisely, MPC protocols



will allow us to transform our FE and HHE schemes from single client to multi-client. Hence, supporting a wide range of use-cases and therefore making our schemes more practical.

**The novelty of this layer is twofold:** On one hand is the design and implementation of two modern encryption techniques (i.e. FE and HHE) that will allow authorised users to efficiently perform computations on encrypted data. On the other hand, HARPOCRATES encryption techniques will be designed in such a way that will allow several ML models to analyse and process encrypted data and eventually output meaningful conclusions without breaching users' privacy. Thus, the HARPOCRATES Crypto layer will promote the intersection of ML with cryptography and has the potential to provide a pathway to groundbreaking technological capabilities.

- L3. ML Layer:** The ML layer will be a core HARPOCRATES layer allowing us to explore a major puzzle: *"How to analyse encrypted data in a privacy-preserving way (i.e. without decrypting them)"*. The ML layer will contain several ML models deployed on commodity cloud infrastructure. The underlying models will take as input encrypted data from the Crypto layer and will be able to classify them with high-accuracy without learning anything about their contents. **The result of the classification will be made available only to authorised parties** (e.g. an analyst sending a legitimate request). Even the cloud service provider running the underlying ML models conducting the classification will not be able to tell anything about the output sent to the analyst. While the main goal of the ML layer will be to classify encrypted data in a privacy-preserving way, it will show how to develop fully privacy-preserving ML models by protecting all phases. More precisely, in the core of the ML layer, apart from the classification algorithms there will be protocols and mechanisms for privacy-preserving feature selection and training. To achieve this, we will utilise the MPC protocols from the Crypto layer and split learning, a collaborative learning approach known. By doing this, we will be able to: (1) rank features in a privacy-preserving manner without the need to pre-process data in a dataset (privacy-preserving feature selection) and (2) protect users' privacy by allowing training without sharing users' raw data to the server that runs an ML model (split learning).

**The novelty of this layer is threefold:** First, the design and implementation of ML models that will be able to classify encrypted data (text, images, video and audio) with high accuracy in an efficient way. Second, the protection of users' privacy during the training phase by incorporating split learning and privacy-preserving feature selection techniques. Third, we will demonstrate how two competing ideas, data analytics and privacy, can be entwined.

- L4. DP Layer:** The DP layer will act on top of the Crypto layer to secure the privacy of the data. In particular, when using FE or HHE to periodically publish statistics, **a well-formed sequence of queries can bypass the confidentiality offered by the Crypto layer** and eventually break the security properties of the underlying schemes. To this end, **all data will be filtered through the DP layer, where additive noise mechanisms will be applied** on them, either before encryption or before publication.

In the case of PPML, the DP layer will be responsible for training datasets in a privacy-preserving way using the following two existing approaches as well as seeking new and more efficient alternatives: (1) knowledge distillation approaches, that require massive quantities of public data (i.e. data that will not receive privacy protections) in addition to massive amounts of sensitive data (that will be protected) - these requirements limit their applicability; (2) adaptations of stochastic gradient descent that are more widely applicable but result in low accuracy compared to non-private models. Noise is added to the gradient estimate and then parameters are updated. The result is a biased and noisy gradient that causes final model accuracy to deteriorate.

**The novelty of this layer is manyfold:** Our DP layer will have three main properties that will make it essential for our ML applications: (1) composability, (2) group privacy, and (3) robustness to auxiliary information. Composability enables modular design of mechanisms: *if all the components of a mechanism are differentially private, then so is their composition*. Group privacy implies graceful degradation of privacy guarantees if datasets contain correlated inputs. Robustness to auxiliary information means that privacy guarantees are not affected by any side



information available to the adversary. Apart from that, we will seek new alternatives to training datasets, by treating the entire dataset as sensitive.

**L5. GDPR Compliance, Legal & Ethics Layer:** The legal and ethics layer is interconnected with the technological layers and pervades the entire HARPOCRATES architecture. It demonstrates how the project will further the state of the art to support diverse GDPR-compliant, privacy-preserving and ethical data spaces for research and digital services in the EU.

This layer will proceed from a systematic analysis of existing knowledge about the potential and limitations of privacy-preserving technologies in automating GDPR compliance and contributing to privacy-preserving and ethical data processing in the EU. We will assess the positive impacts stemming from the development of the project technologies for legal principles under the GDPR, as well as for human rights and ethical values.

Previous projects have explored the limits of formalising, in a machine-readable manner, legal and policy documents to empower data controllers with robust tools for automated compliance. However, most projects focused primarily on achieving complete anonymisation and, did not adequately consider the legal and regulatory barriers to technological adoption, such as the differences in the national implementation of the GDPR across EU member states.

**The novelty of this layer is twofold:** First, HARPOCRATES will advance beyond the state of the art by combining the most advanced encryption and ML techniques to allow users analyse encrypted data, thus minimising the risk of privacy breaches. Beyond the potential of achieving complete anonymisation through privacy-preserving analysis, we consider a broader array of GDPR principles and legal protections. Second, HARPOCRATES will assess the legal challenges introduced by cross-border data sharing to scalability and realistic applicability of privacy-preserving technologies, offering tools and recommendations for data controllers and policy-makers.

**L6. HARPOCRATES Demonstrator Layer:** In order to derive and validate realistic requirements towards the above described five layers, and to assess how the implemented layers can be applied in practice, HARPOCRATES will implement two realistic use cases. These use cases are both representing cross-border data sharing scenarios within federated data infrastructures, coming from the healthcare and the public sector/local government domains, and they both demonstrate how the various tools of HARPOCRATES can be utilised for privacy preserving computation. In both demonstrators multiple HARPOCRATES partners are involved, coming from different countries and institutions, but already working in a collaborative way to share and process data in a privacy preserving manner. Detailed description of the demonstrators is given in the next section.

**The novelty of this layer** is provided by the unique way how the demonstrator applications handle, share and process data in privacy preserving way. Both demonstrators produce large amounts of data (in the order of terabytes), that need to be processed collaboratively using advanced analytics and ML algorithms. Various HARPOCRATES services will support both scenarios, including PPML and FL.

### 1.2.3 HARPOCRATES Demonstrators

To assure that the advanced security solutions developed by the project respond to realistic user needs, HARPOCRATES will implement two demonstrator applications, that are described below.

#### **Demonstrator 1: Threat intelligence generation and sharing between local authorities**

**Partners:** In this demonstrator, two local authorities, one from the Aragon region of Spain and one from the Veneto region of Italy participate. The technical development is supported by S2 Grupo, a Spanish technology company, specialising in Cybersecurity.

**Problem Description:** A threat intelligence sharing scenario between two collaborating European regions (Aragon and Veneto) will be implemented. Governments are one of the main targets for cybercrime attackers and are exposed to great risks, as they provide important services such as health, education and social services. To fight such risks, both regions have significantly large relevant datasets that can be collaboratively exploited using Privacy Preserving Machine Learning (PPML) techniques.



One of the key factors when fighting cybercrime has always been sharing information among different organisations, either public or private. Traditionally, sharing information on cybercriminals new techniques, trends and objectives or even ongoing campaigns has been a common practice in order to prevent attacks or, at least, to detect them at an early stage or mitigate their effects. More recently, the application of Machine Learning, such as Deep Learning algorithms, in cybersecurity has gained significant popularity due to its flexibility and capability to cope, not only with known threats, but also with unknown threats (such as zero-day threats). Since these algorithms require large amounts of data to be trained, their increased popularity has further emphasized the need for data sharing.

However, sharing large amounts of data related to cybersecurity among different entities is often complicated, not only due to the complexity and heterogeneity of the data, but also due to their potential sensitivity. For instance, a neural network could be trained to detect phishing campaigns at a company by inspecting employees' emails and calculating the degree of similarity to existing phishing campaigns. However, that would require the company (and its employees) to give permission to inspect these emails, which may contain private and sensitive information. Also, companies and individuals are reluctant to share data regarding the ways in which they have been attacked, especially when attacks have been successful, because they do not want their public image to be damaged.

In this demonstrator, both participating organisations (the local governments of Aragon and Veneto) will benefit from the advanced data sharing mechanisms provided by the HARPOCRATES framework which let them share private and sensitive information without damaging their public image or infringing any privacy law. Such a mechanism will help collect and exchange new indicators of compromise and threat intelligence, leveraging cybersecurity to a point in which it can cope with the above-mentioned increase in the degree of intelligence and sophistication of cybercrime.

**Planned demonstrator application:** The demonstrator will be implemented as follows:

1. Dataset building: including (1) selection of the subset of users/hosts in each organisation which will be part of the dataset; (2) recovering data during each organisation's normal activity and basic anonymisation to allow for the publication of the datasets to the rest of the consortium; (3) injection of malicious logs for threat detection and sharing experiments.
2. Threat intelligence platform design and architecture: including (1) design of the demonstrator platform, utilising HARPOCRATES services, through which threat intelligence will be anonymised and shared; (2) threat modelling design.
3. ML training using anonymised data, leveraging the PPML services developed by HARPOCRATES.
4. Implementation and evaluation of the demonstrator. The evaluation will include (1) analysing the threat landscape before HARPOCRATES; (2) characterisation of threats and the data required to prevent them; (3) comparison of preventable threats with and without HARPOCRATES services.

#### **Demonstrator 2: Collaborative use of Machine Learning in Sleep Medicine**

**Partners:** Three sleep medicine centers from three different countries (Charite University Hospital in Germany, Kuopio University Hospital in Finland and Hospital de Dieu from France) will be involved in this demonstrator. The technical development will be supported by the medical informatics team from the University Hospital of Gottingen Germany.

**Problem Description:** Machine and deep learning-related research have gained great interest during the last years in the field of sleep medicine. Research projects utilising these methods need to exploit a huge amount of clinically sensitive data. This raises several issues, such as the openness of research, privacy concerns, and security of data, and the importance to find a balance between them. Currently, research projects are supposed to follow the FAIR data principles (findable, accessible, interoperable, and reusable) and the recent European Commission's General Data Protection Regulation (GDPR, 2016/679). However, following the GDPR hinders the implementation of FAIR and makes data sharing between clinical or research institutions highly difficult, especially in the case of retrospective data.

Ideally, data sharing should contain three steps: (1) data encryption, (2) data sharing with secured platforms, and (3) secure data storage. However, data encryption has not received the attention it deserves albeit being one of the key elements: if for any reason the two other steps fail, data encryption





ensures that the data is still protected and identification of individuals is not possible. Thus, having sophisticated and validated data encryption algorithms available would be highly valuable for the whole research community. Additionally, performing complex analytics and machine learning operations on the encrypted data is also desired to facilitate the joint exploitation of this securely shared data.

**The planned demonstrator application:** The demonstrator will utilise data consisting of sleep recordings (i.e., electroencephalography, electrooculography, electromyography, breathing signals, oxime-try signals, snoring (audio), sleeping position, respiratory effort, and possibly the nocturnal video recording), patient medical records and questionnaires related to sleep disorders in text format (word, excel, etc.), collected within the H2020 [SLEEPREVOLUTION](#) project. The aim of the demonstrator is to illustrate that after the encryption the sleep recordings can be handled and analysed similarly (using complex analytics and ML techniques) to the non-encrypted data. It is necessary that the solution can be used regardless of the sleep recording software or device used to analyse or collect the data. Encryption should be possible also regardless of the number or quality of signals in sleep recordings. The solution should include a user-friendly user interface and be readily usable in the clinical (operational) environment by clinical practitioners.

#### 1.2.4 Links with National and International Research and Innovation Activities

The outcomes of related national and international initiatives will be utilised in two different ways. HARPOCRATES partners are involved in several relevant initiatives and will directly apply the outputs of those projects, as listed in [Table 1.2i](#). Additionally, we identified recent EU projects that partially addressed some related areas to HARPOCRATES ([Table 1.2j](#)). The outcomes of these projects will be thoroughly studied, and their open results will be utilised by HARPOCRATES where appropriate.

HARPOCRATES Partners	Initiative	Topic Related to HARPOCRATES	HARPOCRATES WP
UoW, TUNI, RISE, CRT,	ASCLEPIOS	Cryptographic Cloud Storage and Functional Encryption; Deployment of framework components as microservices using MiCADO; sleep medicine prototype.	WP1, WP3 WP4
RISE	CONCORDIA	(1) Develop next-generation cybersecurity solutions and (2) Provide expertise to European policy makers and industry.	WP1–WP6
CBIT	NGI DAPSI	Privacy of personal data in federated data spaces.	WP2, WP3
TUNI, ZEN	IMI2 FACILITATE	Secure data sharing platform (project is expected to start in 2022).	WP1, WP3
S2	CYBERSANE	Identification and standardisation of threat intelligence data, management of forensic artefacts and data.	WP4
CRT	SLEEPREVOLUTION	Sleep apnea data collection.	WP4
S2	SEGRES	Threat modeling for Industrial Control Systems.	WP4
TRI	TRUSTaWARE, D4FLY	Data protection, ethics and human rights impact assessment of privacy-preserving technologies that will feed into the work undertaken by TRI within HARPOCRATES.	WP6

Table 1.2i: Projects directly linked to HARPOCRATES

Project	End Date	DP	FE	HE	HHE	MPC	GDPR	PPML	FL
KONFIDO	2019			✓					
SODA	2019	✓				✓			
PAPAYA	2021	✓	✓	✓		✓	✓		
ASCLEPIOS	2021		✓			✓	✓		
MUSKETEERS	2021	✓		✓		✓		✓	✓
SPHINX	2021			✓					
MOSAICrOWN	2021					✓			

Table 1.2j: Projects with research areas related to HARPOCRATES



### 1.2.5 Interdisciplinary Approach

Improving technologies related to scalable privacy preserving federated computation requires an interdisciplinary approach, involving multiple stakeholders. Therefore, HARPOCRATES brings together experts from a wide range of areas to tackle this issue in a unique way. The core of the work is centred around cybersecurity research, represented by leading European cybersecurity research groups (TUNI, RISE) and SMEs (CBIT, ZEN, S2). Scalable federated cloud computing is another aspect that requires specific attention (UoW, CBIT, ZEN). However, these technical research areas need to be efficiently complemented with expertise in legal and ethical aspects (TRI), assessing and ensuring compliance with data regulations and the GDPR. Finally, the work has to be rooted in users' needs, taking their expectations into consideration. Therefore, HARPOCRATES incorporates experts from the medical field representing four European hospitals in three different countries (CRT, UMG, UEF, VIFASOM), and two local governments from two European regions (SARGA, VENETO), in cross-border scenarios. This unique combination of security, distributed computing, legal, medical and governmental experts provide an excellent basis to develop the next generation of privacy preserving technologies enabling advanced data analytics, Machine Learning and Federated Learning.

### 1.2.6 Gender Analysis and Balance

The Consortium is aware of the regulations of the European Union on gender issues and acknowledges the principle of equality between women and men to both eliminate inequalities and promote equality, as signed in the Treaty on European Union, the Treaty of Amsterdam (May 1, 1999). To this end, gender analysis in the context of HARPOCRATES is examined from a two-fold perspective:

**Gender balance in the specific research and innovation domain:** The research, development and innovation activities carried out in the context of HARPOCRATES can be considered gender and sex neutral. The domains tackled, both from the R&D (data privacy, analytics, privacy-preserving ML and cloud computing technologies) and from the industrial and business communities' perspective, do not provide uneven opportunities, nor they are biased in any way. Nevertheless, to minimise the likelihood of gender or sex specific bias, the project will adopt the policy measures for benchmarking gender equality in science by the European Commission. However, no such bias is anticipated.

**Gender balance in the demonstrators:** Each demonstrator addresses a particular use case where natural differences might exist regarding gender, age, socio-economical level, and other demographic parameters. Discovering and properly addressing such differences is necessary to guarantee equality. Although this specific research topic is outside the scope of this project, the design of the demonstrators will take into account equality of gender and other characteristics to prevent that the results obtained in this project are meaningful only for a privileged population. The project will take the following measures to tackle above issues from both analysed perspectives:

- The technologies, including ML, protocols, software and interfaces, will be developed according to Gendered innovations 2 guidelines<sup>2</sup> to ensure the gender neutrality of such technologies.
- The gender neutrality aspects will be considered during preparation of skills development activities, including online courses and training materials.
- The project will encourage a gender-balanced representation within the consortium, based on equal opportunities and current EU policies.
- HARPOCRATES will encourage participation of female scientists and engineers in open positions during the recruitment campaigns by the partners.
- Most HARPOCRATES partners already have a Gender Equality Plan (GEP) in place. However, all partners are committed to further develop their GEP in their HR policies, including training on gender equality, unconscious biases within their organisation with dedicated resources to implement the plan with close monitoring and period reporting on the progress to the top management.

<sup>2</sup>European Commission (2020). Gendered innovations 2: How Inclusive Analysis Contributes to Research and Innovation.



- HARPOCRATES will ensure that all communication, dissemination, and exploitation actions are using gender-neutral language, stating the project's support for equal opportunities for men and women and non-binary individuals. The analysis of stakeholders and audiences will include how sex and gender of the groups evolved in the different activities, as well as if gender and sex was influenced by factors such as age, type of represented entity and/or location. Results will be reported by disaggregating sex and gender in texts, tables and figures.

### 1.2.7 Open Science Practices

As it is expressed by Objective 6 (Table 1.1f), HARPOCRATES is committed to follow Open Science principles and best practices, and to make its research results reproducible. To achieve this objective, HARPOCRATES will carry out the following concrete actions:

- All security components of the HARPOCRATES Framework developed by academic partners will be **open source**, also facilitating commercial exploitation (by project partners and external collaborators) based on a **dual licensing policy**.
- All technical **deliverables** of the project detailing its research results and outcomes will be **"Public"** and published on the project's website, after acceptance by the European Commission.
- HARPOCRATES acknowledges the importance of **open access** to scientific publications and research data. The project will assure that the EU guidelines about open access scientific publications and research data under Horizon Europe are fulfilled, in line with the strategy developed by the EC.
- Scientific publications will be made available as **Gold Open Access**. Financial resources are allocated in the budget to cover related expenses.
- Final peer-reviewed manuscripts accepted for publication will be **deposited** in an **OpenAire** compliant repository for scientific publications, such as **Zenodo**. Results will be published on the Open Research Europe Platform and the open access repositories of the academic partners.
- Partners will ensure that **metadata** of deposited publications is open under a Creative Common Public Domain Dedication or equivalent, in line with the **FAIR principles** and partners will provide sufficient information (e.g. authors, title, date of publication and venue, funding acknowledgment, project name, acronym and number, licensing terms, persistent publication identifiers, etc.).

### 1.2.8 Research Data Management

Research data, other than publications and source-code as detailed in the previous section, will be generated primarily by the HARPOCRATES demonstrators. The project will develop a Data Management Plan (??) that will detail data management aspects further. A short summary is given below regarding the research data generated by the demonstrators and how the **FAIR principles** are addressed.

Data Management Aspect	Threat Intelligence Generation and Sharing	Collaborative Machine Learning in Sleep Medicine
Types and size	Key events that occurred on users' computers, collected through systems monitoring events. Size: several Terabytes	Sleep recordings (audio/video), medical records, questionnaires. Size: ≈2 Terabytes.
Findability	A coherent and useful metadata model will be defined, publishing the objects and entities that can be uniquely identified, such as typologies of events, software components, etc. The vocabulary will be annotated according to open standards and published through an API.	Each dataset will be described with rich, machine-readable set of metadata, containing context information for the correct interpretation of research data. URN-type Persistent Identifier (PID) will be generated for each file. The collected data is stored to secured servers of the institution in question.





Accessibility	Open standards will be used to share datasets and metadata models. An endpoint will be published offering REST services with JSON for object encoding and HTTPS as transport protocol. The services will be secured by standard means. Authorisation to access datasets with sensitive information will be managed off-line.	Controlled access to datasets will be provided to all consortium partners. After the end of the project, access to the data could be provided to third parties after reasonable request. However, open data sharing can be limited, thus, we must balance between openness, information protection, IPRs, privacy concerns, and security.
Interoperability	Usage of common vocabularies and ontologies will be maximized JSON-LD or similar linking standards will be used. Datasets will be documented and resolvable with globally unique and persistent identifiers.	Metadata will use a formal, well-established machine-readable formalism. We will clearly identify relationships between datasets in the metadata e.g. describing their scientific links to each other.
Reusability	Datasets will be published with rich metadata describing the context where data was generated, such as its scope and date. Datasets with anonymised personal information will be published with a non-restrictive CC-BY license. At the code level, public interest software modules will be published under EU Public License.	The long-term usage of the data and codes will be allowed. After the project, all produced data can be shared to any other user on request under creative commons CC 4.0 license after paying attention to the possible limitations of open data sharing. Codes are stored and shared between the project collaborators through GitHub servers.
Curator Team	<a href="#">SARGA</a> , <a href="#">VENETO</a>	<a href="#">CRT</a> , <a href="#">UMG</a> , <a href="#">UEF</a> and <a href="#">VIFASOM</a>

Table 1.2k: Research Data Management

## 2 Impact

### 2.1 Project's Pathway Towards Impact

HARPOCRATES is aligned with the European strategy for data: it aims to ensure that more data becomes available for the use in the economy and society, through ensuring protection of the personal data with a new federated processing infrastructure, data sharing tools, mechanisms, and architecture. This will help us create a credible pathway contributing to the following impact of the Strategic Plan 2021-2024: "Increased cybersecurity and a more secure online environment". In this section, we specify the nature of our contributions to the outcomes and impacts, i.e., how we categorise our contributions as scientific, economic, technological and/or societal. We also specify the scale and significance of our contributions.

**Expected Outcome: Improved scalable and reliable privacy-preserving technologies for federated processing of personal data and their integration in real-world systems**



 Associated with document Ref. Ares(2022)4525586 - 20/06/2022

***HARPOCRATES contribution to the expected outcome:*** The solution proposed by HARPOCRATES aims to make substantial contributions towards the reliable federated processing of personal data in real-world systems by facilitating the deployment and validation of innovative encryption methods and techniques in realistic demonstration environments – for data processing between stakeholders within complex ICT ecosystems. The contributions will result in the following outcomes:

- Scientific – Design and implementation of ML enablers able to classify encrypted data without decrypting them (text, images, video and audio) with high accuracy in an efficient and privacy preserving environment deployed on commodity cloud infrastructure.
- Scientific – Combining Functional Encryption (FE) and Hybrid Homomorphic Encryption (HHE) to enable efficient processing and classification of encrypted data, beyond the current limitations.
- Economic – Reduced loss from the leaked data for the European public and private sector by applying advanced encryption techniques.
- Economic – Increased number of applications and services due to the resolved security gaps, and as a result, a potential number of new jobs and opportunities open by utilising HARPOCRATES technology.
- Technological – Technology validated in a federated machine learning and data sharing scenario between two European regions and in a cross-border healthcare use case.
- Technological – Consent and interoperability ontology enabler for personal data EDI with enforced GDPR.
- Societal – Increased trust in the public and private services working with the private data, as demonstrated via two realistic demonstrator scenarios.



**Scale and significance of project's contribution:** The HARPOCRATES framework will provide a complete validated solution for enhancing the security, privacy, and data protection of all organisations of the European ecosystem. Our technology will make B2G data sharing a sustainable practice in the EU by providing reliable privacy-enhancing technologies, together with comprehensive data protection approaches to ensure the safe re-use of personal data and commercially confidential business data for innovation and statistical purposes. Performance indicators: 1) [significance: high, scale: Europe-wide] Industry and businesses will have more data available to innovate as a result of the data strategy (at least 50% increase); 2) [significance: medium, scale: demonstrators] Dozen of new jobs and opportunities created as a result of enhanced data sharing practices as an outcome of the demonstrators; 3) [significance: high, scale: regional] At least two European regions are connected and sharing data from 10k users per region; 4) [significance: high, scale: Europe-wide] Minimised economic loss and data leaks for entities employing our enablers. Related Objectives: O3–Table 1.1c, O4–Table 1.1d, O5–Table 1.1e.

**Expected Outcome: More user-friendly solutions for privacy-preserving processing of federated personal data registries by researchers. Improving privacy-preserving technologies for cyber threat intelligence and data sharing solution. Strengthened European ecosystem of open-source developers and researchers of privacy-preserving solutions.**

**HARPOCRATES contribution to the expected outcome:** The impacts will be achieved by making data more widely available by opening up high-value publicly held datasets across institutions, influencing and impacting a variety of stakeholders that will access the system and will use HARPOCRATES technology to process personal data from multiple datasets, finding biases, providing new explanations for diseases, and identifying physical threats from cross-border institutional shared data. One of the HARPOCRATES demonstrators will specifically focus on sharing and utilising cyber threat intelligence data between two collaborating local authorities. To achieve the above impacts, HARPOCRATES will improve existing and will provide new, open source, well documented and user-friendly results with the following outcomes:

- Scientific – Functional encryption schemes in symmetric and asymmetric key settings, supporting a rich set of statistical functions, novel threat-models considering information leakage during the FE scheme's runtime, and design approaches for generic conversion of message-private multi-input FE schemes to function-private schemes.
- Scientific – Hybrid Homomorphic Encryption (HHE) with new symmetric ciphers optimised for machine learning and other large integer use cases.
- Scientific – High accuracy encrypted image classification with PPML, FE, HHE.
- Scientific – Differential Privacy with FE and HHE schemes to build private encrypted databases.
- Scientific – Byzantine-tolerant FL architecture with global data privacy.
- Scientific – Split learning and privacy-preserving feature selection techniques with interfaces and well documented repositories for protection of the data during the training phase.
- Societal – Improved decision and policy making processes and services provided by local governments via shared and improved threat intelligence.
- Economic/technological – Increased business opportunities provided by a set of open-source tools supporting secure data sharing and federated processing.

**Scale and significance of project's contribution:** Our contribution will enable thousands of scientific organisations and technical contributors to create open-source privacy preserving federated data processing enablers. HARPOCRATES will release technology as open services and frameworks, in line with ISA JoinUp tools for public administrations and citizens. Performance indicators: [significance: high, scale: demonstrators] Reference implementation of at least four ML/AI tools and models, using common building blocks, allowing interoperability, and processing personal data: (1) Functional Encryption, (2) Protocol combining DP and FE privacy-preserving statistics, (3) Privacy-Preserving Machine Learning, (4) split learning, (5) GDPR interoperability data sharing tool. Related Objectives: O1–Table 1.1a, O2–Table 1.1b, O3–Table 1.1c, O4–Table 1.1d, O5–Table 1.1e.

**Expected Outcome: Contribution to promotion of GDPR compliant European data spaces for digital services and research (in synergy with topic DATA-01-2021 of Horizon Europe Cluster 4)**



**HARPOCRATES contribution to the expected outcome:** Ensuring compliance with European legal, ethical and privacy principles and regulations is a critical but also ever-evolving process, especially when it relates to data-driven environments, such as the one envisioned by HARPOCRATES. New concepts and ways to handle and process data, which will emerge during the development of the HARPOCRATES platform, may result in difficulties or ambiguities on the ways regulations should be applied. In order for HARPOCRATES to be ready to tackle such issues, a dedicated work package (WP??) is therefore created in charge of ensuring compliance of HARPOCRATES' innovations with EU-wide legal, ethical and privacy principles, including above all the GDPR (EU 2016/67). HARPOCRATES will promote GDPR compliant European data spaces for digital services and research through the following novel contributions:

- Technological, Scientific – Novel combination of the most advanced encryption, ML and DP techniques that will minimise risk of privacy breaches without limiting the ability of authorised users to analyse encrypted data.
- Technological, Economic – Development of automated GDPR compliance tools for data controllers at the fraction of the cost of specialised legal-tech services.
- Technological – Application of this combined novel technology to two cross-border case studies involving both health and other public services data in a number of different European countries.
- Scientific – Comprehensive legal, regulatory and policy analysis of the challenges brought upon by cross-border data sharing scenarios in relation to the GDPR.
- Economic – Reduction of GDPR compliance costs for cross-border data sharing in digital services and research by developing and applying the HARPOCRATES technology.
- Scientific, Societal – Integrated data protection, ethics and human rights impact assessment of the most advanced privacy-preserving technologies.

**Scale and significance of project's contribution:** Thousands of data controllers, including both public sector bodies, SMEs and businesses, will be able to use the HARPOCRATES automated GDPR compliance tools, supported by the combination of the most advanced privacy-preserving technologies and validated on a large amount of patient and citizen data in the project's use cases. Millions of patients and citizens in Europe will benefit from more robust GDPR-compliant protection while facilitating wider use of their data that can act as a catalyst for research, innovation and digital services. The organisations adopting the HARPOCRATES framework will be in a position to achieve compliance with all relevant EU legislation related to security, privacy and data handling (collection, classification, and secondary use). Performance indicators: [significance: high, scale: demonstrators and Europe-wide] 20% reduction in legislation violations reported for HARPOCRATES users compared to the overall ecosystem. Related project objectives: O6 – [Table 1.1f](#), O7 – [Table 1.1g](#)

**Expected Outcome: Strengthened EU cybersecurity capacities and European Union sovereignty in digital technologies. More resilient digital infrastructures, systems and processes. Increased software, hardware and supply chain security. Secured disruptive technologies. Smart and quantifiable security assurance and certification shared across the EU.**

**HARPOCRATES contribution to the expected outcome:** Protective mechanisms and privacy-preserving tools from HARPOCRATES will ensure operational efficiency over the sensitive datasets, while keeping the data protected, thus fostering digital innovation, including cooperation between companies. They will have the ability to deploy their own resilient system using new ISA JoinUP open source HARPOCRATES tools, designed to open a plethora of private data exchanges, that will enhance Europe's strategic autonomy in the digital field. For different verticals, including supply chain, where data exchange was not previously possible, or was problematic as companies gradually lost control over their data, the capacity of the HARPOCRATES innovation is going to provide a resilient digital infrastructure and pathways for new secured disruptive scenarios, not possible before.

**Scale and significance of project's contribution:** Our contributions will complement the growing demand to enhance Europe's strategic autonomy in the digital field: creation of European data spaces to ensure that more data becomes available for use in the economy and society, while keeping companies and individuals in control of their data. The global public cloud market is currently largely dominated by US and Asian companies – HARPOCRATES will help to regain the lack of control over data produced on the European territory kept on non-EU clouds. **Performance indicators:** [significance: high, scale: Europe-wide] Bootstrap of the new European data space for seamless sharing of high-sensitive data for innovative and distributive digital services [significance: high, scale: Europe-wide] Compliance certification scheme for organisations related to distributed computing.

### 2.1.1 Requirements and Potential Barriers

Apart from the strategic impacts outlined by the call, HARPOCRATES aspires to have significant scientific, technological, economic and societal impact, as analyzed above. These impacts are achievable, some



(the short-term ones) within the lifetime of the project, and others (the long-term ones) after mass adoption of the project concepts, tools, technologies and services. Nevertheless, even though these impacts are tangible, they require good will and strong collaboration amongst all stakeholders involved in the specific value chain, and significant effort in order to achieve the changes envisioned. In fact, several barriers possibly hindering the achievement of these impacts should be carefully considered and appropriately tackled. The most important of them are provided in the following text.

**Regulatory Barriers:** Varying EU states regulations and legal frameworks, with special focus on data protection legislation, may cause setbacks in a pan-European adoption of the HARPOCRATES approach. As aforementioned, lack of legal clarity for applications and sensitive information exchange and lack of transparency regarding the utilisation of data collected by such applications and services, as well as inadequate or fragmented legal frameworks are amongst the major barriers hindering the wider adoption of sensitive information exchange and processing. The HARPOCRATES project aims to deliver an innovative cybersecurity framework for efficient processing and classification processes over encrypted data, completely aligned to the directives of the GDPR framework. The aim is to provide privacy-by-design functionalities in an attempt to bring down legal barriers which in turn are vital for deploying (cross-border) services in Europe. This, however, requires pan-European collaboration and tolerance and broadmindedness of the key stakeholders in lowering their own barriers and allowing for the realisation of the proposed change. Non-conformance to EU directives or resilience to adopt the propositions could constitute a major barrier to achieving the project impact. To this end, HARPOCRATES aims to deploy the proposed solution to several key stakeholder categories (organisations, service providers, clients) to highlight the potential and achieve the maximum level of acceptance by all EU member states.

**Ethical Barriers:** While the EU works to fully capitalise on research outcomes, there is a growing awareness of privacy preservation and ensuring that sensitive, personal information is well protected and not exploited and/or misused. Within the context of the HARPOCRATES platform development, all ethical issues will be respected and accounted for (as part of ?? and ??), in order to reduce the barriers that ethical concerns could pose to a wide adoption of the project results in the long term.

**Technological Barriers:** Information technologies develop rapidly, and it is difficult to foresee their evolution, which may influence technical design decisions. This is particularly true in the case of cybersecurity, where new cyber threats of all kinds appear by the minute. Acting proactively so as to stay ahead of the state-of-the-art and deliver a solution that will not become obsolete in the near future, HARPOCRATES will be engaged in a continuous technology watch and safeguard effort assuring that the development process will comply with all related standards, and that new Scientific & Technical requirements may arise will be properly and timely gathered and processed.

**Human Barriers:** We should also be aware of the importance of the human barriers posed towards achieving the suggested impact. The actual scope of the call, aimed at providing scalable privacy-preserving technologies for cross-border federated computation in Europe involving personal data, highlights one of the biggest barriers, namely the human factor and how non-resilience to change by the receivers of the services of the proposed solution, could lead the project to a failure. Towards this end, the HARPOCRATES consortium aspires through the communication and dissemination activities to increase the awareness of, and confidence in the delivered tools, mechanisms and solutions among clients and professionals, and through enabling trust and accountability over personal data to make them active contributors to the service delivery system, increasing their trust in it.

## 2.2 Measures to Maximise Impact

The paragraphs below present a first version of our plan for the dissemination, communication, exploitation, business planning, research data management and IPR management activities. A more detailed plan for the elaboration and implementation of these activities will be presented within the first six months of the project. **Dissemination** activities are focused on the publication of scientific knowledge generated within the context of the project, namely the dissemination of the project's scientific and technological results, mainly through publications and presentations in conferences. Through its



dissemination activities, HARPOCRATES aims at reaching the research community and scientific bodies, including cybersecurity professionals, cybersecurity educational providers and training firms, security incident response teams (CIRT), and computer emergency response teams (CERT) as effectively as possible. **Communication** activities mainly consist of raising awareness about the project through electronic and non-electronic channels and interactive and non-interactive activities, e.g. maintaining a project portal, and presentation in social media. Through communication activities, the project aims at reaching cyber security professionals, as well as the general public. **Exploitation Planning** activities are concerning the exploitation of the project results, whether this is academic, technological or commercial exploitation. Through exploitation activities, the project aims at reaching primarily the healthcare, and public services sectors, as well as the vertical smart city service market associated with it. **Business Planning** activities are complementing the exploitation activities and focusing on the commercial exploitation of the project results and the generation of revenue. The key exploitable assets from the business planning perspective will be the HARPOCRATES framework, its security services and the demonstrator applications. **Research Data Management** activities are aimed to support the data management life cycle for all data that will be collected, processed or generated by the project, as well as the management of publications. Finally, **IPR Management** activities are related to the management of the foreground and background knowledge of the project, complementing and regulating the exploitation and business planning activities.

### 2.2.1 Dissemination and Communication Activities

The dissemination activities, aimed at our stakeholders, will begin early in the project and carry on after the project funding comes to an end. Our communication activities, aimed at the public, the mass media and social media, will take place during the project's lifetime and will include references to project funding from the EU. **Main pillars of dissemination and communication activities:**

- **Objectives:** Informing about the project and its progress as well as to engage the defined target audience.
- **Target Audience(s):** Scientific community, Industry and other relevant EU funded projects (H2020, HEU) and the general public..
- **Channels:** Various channels are utilised (see below).
- **Messages:** These will be adapted to both the target audience and to the specific channel used.

In the following sections we detail the type of activities to be carried out. Main aims are to raise awareness, engage with target audiences and utilise the results to maximise impact. Developed materials will be further disseminated through the target communities to reach a broader audience. Some of these activities can be writing scientific articles, presenting, organising or attending events, issuing press releases, publishing a project website, etc. Project results may be used to develop education and training programs for our tools, as well as to provide developer community instructions.

#### Dissemination of Results per Target Audience

Table 2.2p lists our envisaged target stakeholders and the main messages we intend to convey to them. The first column identifies the stakeholders, second column lists dissemination activities to reach them.

Target audience /Stakeholders	Tools and channels
<b>Academia</b>	<ul style="list-style-type: none"> <li>• Liaison and collaboration with related research initiatives and research groups.</li> <li>• Peer reviewed publications in scientific journals.</li> <li>• Participation in relevant conferences/workshops with a large audience among the scientific community.</li> </ul>
<b>Industry</b>	<ul style="list-style-type: none"> <li>• Liaison with industrial associations.</li> <li>• Key articles in trade press.</li> <li>• Participation in relevant conferences/workshops with large industry presence.</li> <li>• Organisation of workshops and demos.</li> <li>• Individual presentations/discussions with key organisations.</li> </ul>
<b>Public authorities, healthcare institutions</b>	<ul style="list-style-type: none"> <li>• Articles in trade press.</li> <li>• Participation in relevant conferences/workshops.</li> <li>• Meetings with policy makers and representatives of healthcare ministries.</li> <li>• Organisation of workshops and demo events at pilot sites.</li> </ul>





<b>General public / citizens</b>	<ul style="list-style-type: none"> <li>• Communications through social media.</li> <li>• Press releases on project outcomes in national media.</li> <li>• Creation of easy-to-understand videos, published on project website.</li> </ul>
----------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Table 2.2p: Dissemination, Targeted Audiences and Partners

### Dissemination Plan

The dissemination plan describes the measures and channels – detailed in the table below - to build effective awareness of the project results, creating an understanding and aiming for action among the target groups. The execution of this strategy will facilitate the best use and uptake of the outcomes and research insights generated throughout the project's lifetime, reinforcing the targeted impacts.

**Dissemination tools and channels:** We will use the following tools and activities for disseminating our results to our stakeholders, as well as communicating with the traditional media and the general public about the project and emphasising the EU support and the public interest behind our goals. For each tool or activity, we provide a brief description and performance indicators. The dissemination activities are key to the exploitation of results. Our dissemination activities will ensure stakeholders are aware of our innovations and will help prepare the market for our outcomes.

Activity	Description	Performance Indicators
<b>HARPOCRATES Website</b>	Information about the project, the consortium members, the deliverables, project news, a blog, information about the other projects with which HARPOCRATES will cluster. The site will contain attributes such as a mobile-friendly design, quick loading times, accessibility, web analytics, traffic tracking, search engine optimisation and integration with relevant social media.	Live by <u>Month 3</u> , No. of content updates: <u>100</u> , Targeted No. of visits: <2,000: mediocre; 2,000-4,000: good >5,000: excellent.
<b>Guidelines, training materials, flyers and brochures</b>	The partners will develop various training materials (PowerPoint slides and briefing papers), generic flyers and brochures to streamline the transfer of technical knowledge to stakeholders.	<u>two</u> brochures, <u>one</u> flyer, slide sets, a briefing paper (translated into the languages of the countries where we conduct pilots), three posters, No. of downloads of briefing paper <300: poor; 300-1,000: good >1,000: excellent.
<b>Peer-reviewed publications, policy briefs, position statements</b>	HARPOCRATES will organise special issues in top peer-reviewed journals (ACM, Elsevier, Springer, IEEE). A book chapter will be organised titled "Federated data space design and technologies". Peer-reviewed publications in top refereed scientific journals and conferences relevant to the research, as well as publications in technical journals will be targeted.	No of special issues: 2, No of journal publications: 5, No of conference proceedings: 15, Book chapter: 1
<b>Non-scientific Publications</b>	HARPOCRATES will release and contribute to blog posts, articles, books and any other non-scientific publication oriented to end-users and the wider public. Reference media to consider are <a href="#">CORDIS Research*EU Magazine</a>	Non-scientific publications: 20
<b>MOOC, Videos, Blogs</b>	HARPOCRATES will produce Massive Open Online Courses (MOOC), videos will be uploaded to YouTube and/or other social media platforms such as Vimeo, Dailymotion or other similar video platforms. Blog entries regarding technical achievements will be written and published on the website.	Min. one MOOC. No. of blogs on the website: >10. Three videos, the first at M10, second at M20, third at M30. The videos will be about 90 seconds long.
<b>Webinars</b>	HARPOCRATES will produce webinars aimed at different stakeholders. Webinars will be one hour long – half an hour for presentation, and half an hour for questions and discussion.	No. of webinars: at least 5



<b>Conferences and workshops</b>	Partners will showcase project results, present talks, create exhibition spaces and provide personal engagement. A final international cyber security symposium will be organised to present the results and outcomes of the whole project.	Symposium organisation: 1, No. of conferences/workshops attended: 10
<b>Newsletters, email campaigns</b>	HARPOCRATES will use email campaigns and publish a regular newsletter to reach out to end users, key technology communities, and engage with academia.	Min. 36 newsletters, approx. 18 emails (bimonthly) to the project’s stakeholders on the contact list
<b>Open access repositories to research</b>	HARPOCRATES will facilitate an online repository to manage and publish material with different access permissions, including peer-reviewed publications, shareable scientific research data, and other types of resources generated. The project will rely on <a href="#">Zenodo</a> –part of the European Open Science Cloud (EOSC)– to deposit both publications and datasets. In addition, the use of <a href="#">Open Research Europe</a> will be encouraged as an open access publishing platform for scientific articles among the consortium, with no author fees and compliant with open access requirements. Scientific publications will also be deposited in the open access repositories of the academic project partners.	Open access to all on two EU clouds, the project website and the open access repositories of the project partners.

Table 2.2q: Dissemination Plan

**Communication Plan**

Communication activities involve specific measures for publicly promoting the project and its results during its lifetime. Campaigns will be designed and implemented to build traction among the target audience efficiently, capitalising on the multi-language nature of the consortium to cover an international footprint. Additionally, a number of communities and multiplier organisations in various related domains will be targeted by leveraging partners’ networks. The initial list of such organisations will be evaluated and updated later to ensure that the most relevant ones at the given moment are selected. The initial selection will be done at the very beginning of the project (by M3). During the remaining duration of the task, each selected initiative/community will be contacted and collaboration will be initiated to identify how the members of these communities can best benefit from HARPOCRATES. Based on these activities feedback will be provided to WP??-?? in order to adapt/expand the solutions accordingly. We will use existing ecosystems to facilitate such communication activities, for example:

1. FIWARE – several SMEs, entrepreneurs and large companies; our enablers will be adapted to be used as FIWARE specific enablers (e.g. Keyrock identity Management and Wilma Proxy). Partners has worked with FIWARE or contributed to the FIWARE community before (e.g. [UoW](#) is partner in the DIGITbrain project where FIWARE is one of the main supported ecosystems).
2. Confidential Computing Consortium (CCC) is a project community at the Linux Foundation dedicated to defining and accelerating the adoption of confidential computing.
3. OSF Edge - Openstack Edge Computing Working Group - 12,000 people from 130 different countries and yearly Open infrastructure summit events; dissemination activities related to distributed challenges (whitepapers, blogs, position papers, etc).
4. GAIA-X – a pan-European association that develops and promotes federation of cloud services within existing cloud infrastructures. Its activities support open-source software development, data sovereignty and data portability - to communicate its results in the area of privacy-preserving computation in the working groups “Provider” and “Federation Services/Open-source services”.
5. OpenMind – open-source community whose goal is to make the world more privacy-preserving by lowering the barrier-to-entry to private AI technologies.

[Table 2.2r](#) summarises the major activities and performance indicators of the communication plan.





Activity	Description	Performance Indicators
<b>Social media</b>	The partners will use Twitter, LinkedIn, Facebook, Instagram and other social media channels to reach out to the public	2-4 tweets per week, 1-2 Facebook and Instagram posts per week, LinkedIn group established
<b>Podcasts</b>	The partners will develop 10-15 minutes long podcasts summarising the major achievements of the project	5 podcasts
<b>Branding</b>	HARPOCRATES will adopt a common graphic approach throughout the project	Logo and templates for slides, press releases, newsletters
<b>Newspaper stories and press releases</b>	HARPOCRATES will identify and target top journalists and bloggers in each Member State who have written previously about relevant topics to the project.	At least the top five journalists in 10 Member States are identified and approached (5 x 10 = 50)
<b>Translations</b>	The partners will translate press releases, the newsletters and selected other materials into at least the languages of the partners.	At least 15 translations of press releases, newsletters, brochures and flyers

Table 2.2r: Communication Plan

### Exploitation Strategy

The exploitation strategy, that will be part of the “Communication, Dissemination and Exploitation Plan” prepared by M6 of the project, will follow the evolution of the scientific and technological outcomes and will be based on the collective and individual exploitation strategies and activities. The Consortium Agreement will define the IPR management strategy that will also significantly influence the exploitation strategy. The final version of the exploitation plan will produce a reliable strategy for sustaining the partners’ results beyond the end of the project.

**Tentative Business Model and Business Plan:** The HARPOCRATES consortium has already drafted a Business Plan for the commercial exploitation of the project’s outcomes. At proposal stage, it is actually an initial approach, while the detailed plan will be defined within the context of the corresponding work package (WP??) and will depend on actual project activities. This initial approach covers plans of a community and user sharing edition of the HARPOCRATES platform, which will be partly open source, as well as the plan of charging for professional services through which revenue will be raised. The HARPOCRATES consortium favours open-source policies, and it has decided to take a dual-licensing approach towards managing the knowledge produced internally but also exposing specific knowledge to the scientific community. Therefore, the indicative business plan is structured around a two-core business axes: a) professional-oriented services that will be supported by the HARPOCRATES core framework, and b) audience-centric services, which will enable generation and sharing of new tools and experiences. [Table 2.2s](#) below outlines an indicative Business Model Canvas. Its goal is to provide a systematic on-going process by which business “steps” can be reported and monitored, and can iteratively evolve until HARPOCRATES achieves a certain market outreach maturity. Its value is that it serves as a dynamic business plan observatory and scoreboard, which will be updated to indicate whether business potential is on track. Since the development of a business model description requires extensive knowledge of the business parameters and actors, including the customers and the most technical aspects of what an entity does, partners will jointly work on such a task. Preparing a complete business model description will require the contributions of entities outside the consortium and the overall ecosystem.



<p><b>Key Partners:</b> Critical ICT infrastructures; academia, ICT vendors and asset providers; open source communities; standardisation organisations; investors; early adopters and promoters.</p>	<p><b>Key Activities:</b> Development of cybersecurity risk management and services; business model development for new services; maintenance and integration.</p> <p><b>Key resources:</b> Research results from project activities; successful validation of use cases.</p>	<p><b>Value Proposition:</b> Platform for enhancing cyber security, privacy and data protection of complex ICT infrastructures; Secure and private sensitive data exchange with federated learning between organisations.</p>	<p><b>Customer Relationship:</b> Long term support and future upgrades; training support; community support; dedicated assistance.</p> <p><b>Channels Website, social media;</b> demonstration &amp; dissemination activities; standardisation, open source; clustering; fairs and events.</p>	<p><b>Customer Segment:</b> Professionals; organisations administrators; executives and managers of service providers; third party developers; CERTs/CSIRTs; cybersecurity educational providers and training firms; cybersecurity researchers.</p>
<p><b>Cost structure:</b> R&amp;D costs; infrastructure costs (i.e. platform hosting); data centre operations; integration costs; marketing costs &amp; ads; customer support.</p>			<p><b>Revenue streams:</b> Open source versus closed system services (licensing); tools on demand (API); direct platform sales; after-sales service; engineering services; consulting services; development contracts.</p>	

Table 2.2s: Exploitation Strategy

**Individual Exploitation Plans:** Exploitation will concentrate on the commercial partners/SMEs of the project. Academic project partners will also provide value added services based on the result.

Partner	Type of Result
<b>S2</b>	Development of approaches for threat prevention and detection based on AI algorithms for advanced encryption and anonymisation of sensitive and private data. Results are expected to contribute to a 15% growth of the company over the next four years.
<b>CBIT</b>	Exploitation of the project’s results by integrating them into its commercially available ConfidentialCloud confidential data collaboration and portability platform that allows competing entities to engage in trustworthy data collaboration based on verifiable attestation tokens.
<b>SARGA</b>	Knowledge and prototypes developed in the use case will be used to improve the cybersecurity of the organisation, thanks to the integration of the system with the rest of the security tools in the infrastructure. Use case will be shared with the Spanish cybersecurity authorities (CCN-CERT), for further dissemination and implementation among the rest of the Spanish and European regions and governments.
<b>VENETO REGION</b>	Existing cooperation with local industry to introduce the project’s results and contribute to their marketing strategy. In agreement with regional stakeholders and strategic players, involve SMEs and start-ups to raise knowledge and better attitude to exploit the project’s results according to the agreements and follow-up plans established.
<b>ZEN</b>	Integrate the resulting components and modules into the company’s COVIDAS data-driven solution for monitoring patients’ response and potential complications of Antibody-Dependent Enhancement (ADE) in vaccine recipients. Enable collaborative research to allow privacy-preserving data sharing and federated research on multiple healthcare datasets.
<b>UoW</b>	Offering commercial consultancy to commercial entities interested in exploiting results; providing code maintenance, development and customisation on demand. Software developed by UoW will be released under Apache 2.0 open-source licenses.
<b>TUNI</b>	Existing cooperation with industry to introduce the project’s results to key industry players and the large number of startups hosted in the University. This will bring in a large number of master students that will have the opportunity to work on project-related activities as part of their thesis and can carry this knowledge to their employers after graduation.



<b>RISE</b>	Open-source code. Collaborates with industry and will explore joint exploitation of results. Project outputs can lead to deeper studies and follow-up works with exploitation potential.
<b>TRI</b>	Advance ethical, privacy, and data protection practices and services around emerging technologies, as well as to further refine and implement data protection impact assessment methodology. This project will enable TRI to acquire knowledge, capacities and experience that will give advantage in relation to competition for the design and delivery of services.
<b>UEF</b>	Implement the results in clinical and research practice. UEF has a strong and wide collaboration network with top-level hospitals and research institutions worldwide, and thus, actions related to secure data transfer play a key role in their research.
<b>CRT</b>	Exploitation by contributing to clinical studies with pharmaceutical companies requiring encrypted sleep recordings. Implementation for clinical service in sleep center quality control.
<b>UMG</b>	Implementing a national service for encrypted sleep recording including expert annotations for medical quality control for sleep center certification.
<b>VIFASOM</b>	Service implementation for remote sleep recording analysis and scoring as central reading in sleep medicine clinical trials for new drugs and recording devices.

Table 2.2t: Individual Exploitation Plan

### Knowledge Management Strategy and IPR

HARPOCRATES will generate a range of data and knowledge, some of which will be publicly available, some proprietary and some will be for dissemination and communication to the public. At this point, the consortium has agreed to the following principles for handling intellectual property rights within HARPOCRATES: A joint Consortium Agreement, to be signed by all parties at the latest before the Contract's signature. The CA will address (i) the confidentiality of the information disclosed by partners during the project, (ii) ownership of results resulting from the execution of the project, (iii) legal protection of results deriving from the execution of the project through patent rights, (iv) commercial utilisation of results, also taking into account joint ownership of the results, (v) patents, know-how and information related to the use of knowledge owned by one of the partners, resulting from work carried out prior to the agreement, and (vi) sub-licences to third parties within clearly defined limits.

To ensure that these goals are achieved and the results of the project can be efficiently exploited after its completion, knowledge management strategy and IPR-related issues will be continuously monitored and actively discussed during project's lifetime. Further agreements, between all or a smaller targeted subset of project partners will be considered, if and when necessary. Such agreements can go beyond the CA and aim to assure long-term sustainability via commercial exploitation. All project partners will contribute to IPR and knowledge management planning and discussion, taking place in WP???

### Contribution to Standardisation

HARPOCRATES partners will contribute to the work of several standardisation bodies to assure that the results of the project are compatible and interoperable with current and upcoming industrial and professional standards. The main planned contributions to standardisation are summarised below.

Body/ WG/ Con- tribution	Scope
<b>IETF/ NISEC/ TEEP / RATS</b>	TUNI and RISE are active in IETF (Internet Engineering Task Force) and contribute to standardisation on how to build standards for cryptographic cloud storage and analyzing data in a privacy-preserving way. CBIT contributes to the IETF Trusted Execution Environment Platforms (TEEP) and IETF Remote Attestation procedureS (RATS) working groups, and will promote and reflect the project's results in these forums.



<b>OASIS TOSCA</b>	<b>UoW</b> is a member of the OASIS TOSCA (Topology and Orchestration Specification for Cloud Applications) working group. The MiCADO cloud orchestrator, that is proposed to be utilised in HARPOCRATES for the cloud deployment and management of the demonstrator applications, is compliant with the TOSCA standard specification.
<b>ISO/IEC</b>	<b>UEF's</b> Sleep Technology and Analytics Research Group is a part of the European Sleep Research Society's sleep lab network. The <b>UEF</b> SmartSleep Lab will be a certified and accredited sleep laboratory meeting the international quality, impartiality, and compliance requirements (ISO / IEC 17025, ISO / IEC 17011). Among others, ISO/IEC 27001/2 will be analysed by <b>ZEN</b> , to propose recommendations for additional compliance certification of organisations related to distributed computing.
<b>CEN TC251, ISO TC 215</b>	<b>CRT</b> is member of CEN TC251 and ISO TC 215 related to medical device communication, as well as DIN NaMed (German Institute for Norms and Standards) at the Workgroup on Medical Informatics. The standards for medical device communication and medical software which includes the electronic health care record will be proposed through these WGs.
<b>Confidential Computing</b>	<b>CBIT</b> is a member of the Confidential Computing Consortium (CCC, confidentialcomputing.io) where it contributes to defining and steering vendor technology support for privacy enhancing technologies.
<b>OSF Openstack</b>	<b>ZEN</b> is a member of the OSF Openstack Edge Computing group and will contribute to the pre-standardisation discussion in regular WG meetings coordinated by Openstack.
<b>ISO/IEC JTC 1/SC42</b>	<b>TRI</b> is an active member of ISO/IEC JTC 1/SC42 - Artificial Intelligence, ISO/IEC JTC 1/SC27 - Information security, cybersecurity and privacy protection, ISO/IEC JTC 1/SC 24 - Computer graphics, image processing and environmental data representation, ISO/TC 307 - Blockchain and Distributed Ledger Technologies. Through these committees, contribute to standardisation activities by working with bodies at national, regional and European level.
<b>W3C</b>	Contribution to the W3C GDPR consent ontology for data access - adaptation of the W3C Data Protection Ontology for describing obligations based on a draft version of GDPR ( <b>ZEN</b> and <b>TRI</b> ).

Table 2.2u: Individual Exploitation Plan

### 2.3 Summary

Specific Needs	Expected Results	D+E+C Measures
----------------	------------------	----------------



<p>– Empower the use of <b>big data and ML/advanced analytics</b> from silos of isolated private data across borders and different administrative domains for research and digital services.</p> <p>– Adequate <b>protection of data</b> according to the GDPR can prevent its full utilisation for society. Assure the required levels of privacy in the data (via advanced encryption techniques) <b>without compromising its exploitation</b> for advanced analysis.</p> <p>– Ensure applicability of advanced privacy-preserving computation techniques in real-world use case scenarios.</p> <p>– Scalability and reliability of privacy-preserving technologies in realistic problems.</p>	<p>– <b>A toolkit to support PPML and PPFL:</b> A toolkit to support PPML and PPFL: A set of open-source advanced security services based on a combination of improved FE, DP and HHE mechanisms, with the aim to support Privacy Preserving Machine and Federated Learning and advanced data analytics.</p> <p>– <b>Algorithmic model:</b> Replicable federated cryptographic machine learning conventions and algorithms for identifying and classifying encrypted dataset (text, image, videos).</p> <p>– <b>Privacy preserving cross-border federated data sharing:</b> The demonstration of these technical achievements in two real-world application scenarios supporting cross-border ML and data analytics on federated datasets related to sleep medicine and threat intelligence at local authorities.</p>	<p>– <b>Standardisation &amp; certification:</b></p> <ul style="list-style-type: none"> <li>• Contribution to standardisation bodies IETF, ISO/IEC, W3C, CEN.</li> <li>• ISO/IEC compliance certification of organisations related to distributed computing.</li> </ul> <p>– <b>Exploitation:</b> HARPOCRATES will spin off commercial products and services, led by the SMEs involved in its use cases and tool creation. Moreover, it will deploy a federated privacy preserving machine learning tools in two European regions and one healthcare use case, enabling it to claim its position as a provider of GDPR compliant interoperable tools for data sharing.</p> <p>– <b>Dissemination and communication:</b> Through online and offline events, and using various channels: HARPOCRATES website, wiki, materials, flyers and brochures, peer-reviewed publications and special issues, non-scientific Publications, Massive On-line Open Courses (MOOC), videos, webinars, conferences and workshops, newsletters, mail campaigns. social media, podcasts, branding, newspaper stories and blog posts, translations.</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Target Groups	Outcomes	Impacts
<p><b>Public ecosystem:</b> Public institutions and authorities, risk managers, municipalities, regions</p> <p><b>Industrial Ecosystem:</b> ICT vendors and asset providers; AI Engineers, service providers; third party developers; CERTs/CSIRTs; cybersecurity educational providers and training firms, vertical industries, Network Infrastructure/Service Providers, System Integrators.</p>	<p>– <b>Take-up by public authorities:</b> Adoption of innovative and privacy preserving technology in two European regions processing data of more than 20K users.</p> <p>– <b>Take-up by hospitals:</b> Four European hospital in three different countries using HARPOCRATES tools demonstrating the functionalities of the framework, including federated machine learning capabilities over encrypted datasets.</p>	<p><b>Economic:</b> Industry and businesses will have more data available to innovate as a result of the data strategy (at least 50% increase). New jobs and opportunities created as a result of improved data sharing practices. Reduced loss from the leaked data for the European public and private sector, increased number of digital services working with federated data, and a potential number of new jobs and opportunities open by use of the HARPOCRATES technology.</p>



<p><b>Academia:</b> Researchers and data scientists, private and public research institutions, and scientific organisations</p> <p><b>Social Innovation Sector:</b> Privacy, Security, Legal Agencies, Ethical Researchers, Digital Education Providers, Civil Society Organisations</p> <p><b>Policymakers:</b> European Commission, Regulators, Public Agencies, Observatories/Think Tanks</p> <p><b>Society:</b> General public, Non-specialised media, individuals, society and economy as a whole</p>	<p>– <b>Academia:</b> Initiation of complimentary and competitive scientific research activities including further scrutiny of raw data and technologies - high use of private datasets containing personal data. Six academic partners in the consortium are doing research based on the outcomes, and at least three times more research partners are involved by the end of the project from the associated partners and the larger research community, as a result of the dissemination activities. Successful scientific breakthrough is achieved in the fields of DP, HHE, FE, federated learning and PPML.</p>	<p>– <b>Scientific:</b> Hundreds of scientific organisations, technical service providers and developers deploy open source privacy preserving federated data processing enablers released as open frameworks and ISA JoinUp tools for public administrations and citizens.</p> <p>– <b>Societal:</b> Increased trust in the public and private services working with private data; improved decision and policy making processes and services provided by local governments via shared and improved threat intelligence; better and more reliable healthcare services underpinned by advanced privacy preserving data analytics solutions.</p>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

### 3 Quality and Efficiency of the Implementation

#### 3.1 Work Plan and Resources

The HARPOCRATES project aims to change the way modern online systems and services work by mainly adopting a new broad vision of encryption schemes and privacy-preserving mechanisms. To achieve its objectives, the project will be organised in seven work-packages.

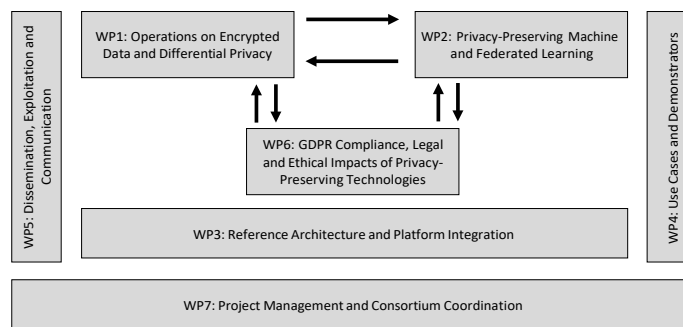


Figure 3: Relations and interactions among workpackages

#### Subcontracting costs



Participant	Cost (€)	Justification
Veneto (10)	30,500.00	To facilitate the technological data exchange with the various stakeholders involved in the project, as well as to support, interpret, analyze and evaluate the evolution of the project requirements and limits to the context of the Veneto Region.

Table 3.1a: Subcontracting

### 3.2 Capacity of Participants and Consortium as a Whole

#### 3.2.1 Consortium as a whole

HARPOCRATES consortium combines multidisciplinary competences and resources from academia, the healthcare, public and law sectors, industry and research communities focusing on applied cryptography, Machine Learning, privacy, cloud security, access control, threat intelligence, GDPR-compliant, privacy-preserving and ethical data spaces for research and digital services in the EU, large scale distributed computing and big data analytics. It consists of thirteen (13) partners from seven EU member states (i.e. Finland, Germany, Greece, Ireland, Italy, Spain, Sweden) and two further European countries (i.e. United Kingdom, Serbia). This multinational as well as multidisciplinary cooperation is essential in order to implement the rising security and privacy requirements of modern privacy-preserving systems, addressing the need for the design and use of services that will respect users privacy. This will allow users to fully control their data while at the same time providing certain guarantees about the overall confidentiality of their private information and the unbiased decisions of the underlying services through the use of privacy-preserving machine learning algorithms.

The project brings together experienced partners with cutting-edge expertise, dedicated to contributing to the successful implementation of the project objectives. The consortium combines **universities, research centres, SMEs, legal experts, healthcare providers and local government organizations** to both make significant advances in research on privacy enabling technologies for cross-border federated computation, and to prove that these advances are directly applicable in real-world scenarios.

The **core research and development** within the project will be carried out by three academic partners (universities and research institutes – **TUNI**, **RISE**, **UoW**) and two R&D SMEs (**ZEN**, **CBIT**) with appropriate capacity in both research and technical development activities. These HARPOCRATES partners represent significant experience in applied cryptography, cyber-attacks surface identification and vulnerability analysis, platform and network security management, cyber threats, Machine Learning, privacy in widely deployed communication networks, cloud security, high-performance and distributed computing, and access control and virtual collaboration environments for clinical research.

Concerning the academic partners, **TUNI**, as one of the leading Universities in Finland, is bringing internationally visible research expertise in applied cryptography, cybersecurity and privacy. **TUNI** has expertise in areas spanning from the theoretical foundations of cryptography to the design and implementation of leading edge efficient and secure communication protocols, privacy-enhancing technologies, and security and privacy in cloud computing with a focus on trusted computing and secure storage. **RISE**, having the largest security lab in Sweden, complements this expertise with significant experience and background in privacy-preserving ML, provable security, trusted execution environments, IoT security, hardware security as well as with extensive experience in industry system and security standardization (3GPP, OMA, XACML, IETF, Trusted Computing, Bluetooth, TCG and ONVIF). Finally, the Centre for Parallel Computing from **UoW**, one of the most diverse and multicultural universities in the United Kingdom, offers research and development expertise and a long standing track-record in the field of secure high-performance, parallel and distributed computing, especially computation in the cloud to edge computing continuum. While **RISE** and **TUNI** will lead the security related research activities (WP1/WP2), **UoW** will coordinate the development and deployment of the demonstrators in the multi-cloud HARPOCRATES testbed (WP4) and will contribute significantly to framework development (WP3). **UoW** and **RISE** will also provide the necessary compute infrastructure in the form





of their private cloud computing clusters (complemented with commercial resources) to develop and deploy the demonstrator applications.

To assure that the results and outcomes of HARPOCRATES are developed to the highest possible standard and will be applicable directly for industry, two SMEs complement the work of the academic institutions. **CBIT** is a Swedish SME, specializing in cybersecurity, cloud and confidential computing. With a focus on deploying and orchestrating applications in confidential computing enclaves, **CBIT** is an active partner in the Gaia-X Sweden hub, working towards a secure and sovereign European data digital infrastructure. Furthermore, it is an active contributor to standardization in IETF. **CBIT** will focus on Privacy-Preserving Federated Learning (WP2) and will lead the development and framework integration activities (WP3). **ZEN** is a Serbian company focusing on research and development activities related to ML, IoT, augmented reality and blockchain technologies. Besides technical development activities, the main contribution of **ZEN** will be related to the exploitation and dissemination of the results (WP5) utilizing their extended network and commercial expertise.

**Compliance with data regulations and the GDPR** is an important complement to the technical knowledge created in the project. HARPOCRATES involves **TRI** as experts in legal, ethical and GDPR related issues. **TRI** is an Irish SME comprised of a team of social and data scientists, ethical, legal and human rights experts who work across the technology-social disciplinary divide. **TRI** will be responsible for coordinating all legal, ethical, human rights and GDPR related activities in the project (WP6).

HARPOCRATES also assures that the developed technological solutions address **users needs** and validated in **realistic, real-word federated data infrastructures**. To provide such settings, HARPOCRATES implements two demonstrators involving partners from different EU countries in cross-border federated data sharing scenarios. **Threat intelligence generation, sharing and utilization** via Privacy-Preserving Machine Learning is implemented by a team involving two local governments, one from the **Veneto region of Italy (VENETO)** and one from the **Aragon Region of Spain (SARGA)**. These organisations have access to the required data and already work in collaboration to share important information in order to improve the quality of their services. Technical expertise to implement the demonstrator will be provided by **S2**, a Spanish technology SME specializing in cybersecurity.

The second demonstrator will concentrate on cross-border collaborative use of Machine Learning in sleep medicine. This use case builds on the results of the EU funded H2020 SLEEPREVOLUTION project which focuses on data collection and processing of sleep apnea data in Europe. Three sleep healthcare centers from three European academic hospitals from three different countries (**CRT** from Germany, **UEF** from Finland and **VIFASOM** from France), each of them involved and already collaborating in the SLEEPREVOLUTION project, will contribute with data and related problems/analysis to this demonstrator. To provide the necessary expertise in the design and development of the technical solution, the medical informatics team from the **University Hospital of Gottingen (UMG)** will complement this team and will coordinate and conduct the implementation of the demonstrator.

Finally, to efficiently manage a project of such scale, significant operational capacity, expertise and experience in project management is required. As a large university, **UoW** has such operational capacity to manage the HARPOCRATES project. Moreover, the **UoW** team behind HARPOCRATES has been successfully coordinating and managing several European research projects in the past, the most recent of these being the H2020 **COLA** and **ASCLEPIOS** projects. The successful project coordinator of **COLA** and **ASCLEPIOS**, **Prof. Tamas Kiss**, will fulfil the same role in the HARPOCRATES project too, assuring the organized and smooth running and coordination of all activities. Moreover, **UoW** will be supported by **TUNI** and its principal investigator **Prof. Antonis Michalas** who will act as scientific coordinator, overseeing and assuring the high quality of the core research activities.

As at the time of signing the Grant Agreement the UK's affiliation to Horizon Europe has not been finalized, the role of the coordinator has been shifted from **UoW** to **TUNI**. All formal coordination that requires interaction with the EC will be carried out by **TUNI**.



## 4 Ethics self-assessment

### 4.1 Ethical dimensions of the objectives, methodology and likely impact

The main ethical dimensions of HARPOCRATES relate to the collection and processing of personal data for the purposes of producing novel insights in the areas of Privacy Preserving Machine Learning (PPML) and Federated Learning (FL), based on cutting edge and further enhanced cryptographic solutions, and in-line with GDPR.

All personal data collected during the project will be kept secure and unreachable by unauthorised persons. The data will be handled with appropriate confidentiality and technical security, as required by law in the individual countries and EU laws and recommendations. The protection of personal data will be ensured through procedures and appropriate technologies and appropriate European and Internet security standards from ISO, ITU, W3C, IETF and ETSI. To assure the participants privacy, all data will be anonymised or pseudonymised, encrypted and stored in appropriate files to which only the relevant staff have access. Personal information will be stored and used up to five years after payment of the last project payment or until consent is withdrawn (the sooner applies), after which point it will be deleted. All personal data will be deleted from the databases of the consortium with proper software and/or hardware procedures rendering unauthorised restoration impossible. In case any partner has a parallel, legal basis obligation to further process any kind of data, those partners will further process the data on the given legal basis, while all other partners shall carry out the deletion process. The details on storage, process and deletion will be contained in the project's Data Management Plan.

Additionally, the HARPOCRATES demonstrations will feature the deployment and use of artificial intelligence, and, more specifically privacy-preserving machine learning (ML) techniques which will be applied both to encrypted and raw data. Ethics issue that may raise in this case include human agency and oversight of data-driven technologies, privacy, transparency, fairness, diversity and non-discrimination and accountability. Our approach to addressing these issues relies on both technology design choices and our integrated impact assessment methodology. First, we will achieve high accuracy in classification of encrypted data outperforming existing approaches, thus minimising the extent to which our AI models will relate to and affect human subjects. Second, we will develop a comprehensive ethics, data protection and human rights impact assessment to systematically monitor ethics issues relating to the use of AI in HARPOCRATES and implement effective mitigation strategies throughout the project.

All partners in the consortium will adopt good practice data security procedures. This will help avoid unforeseen, usage or disclosure of data, including the mosaic effect, i.e., obtaining identification by merging multiple sources. Measures to protect data will include access controls via secure logins, installation of up-to-date security software on devices, regular data back-ups, etc. Recorded information (audio and/or visual) will be given special consideration to ensure that privacy and personal identities are protected. Furthermore, the HARPOCRATES consortium will adhere to the principle of data minimisation, as well as carefully consider anonymisation and pseudonymisation safeguards. With regard to anonymisation and pseudonymisation, HARPOCRATES will adopt a 'anonymise if possible, then pseudonymise if possible, then only use personal data if strictly necessary' policy.

To support data subject access rights, HARPOCRATES will provide data subjects the following information:

- Identity and contact details of HARPOCRATES data controller(s).



- Purpose of the data processing and its legal basis.
- Legitimate interests of the controller(s) and third parties.
- Categories of personal data and their sources.
- Recipients or categories of recipients of personal data.
- Transfers to third countries and safeguards, as applicable.
- Retention periods and relevant criteria.
- Data subjects' rights: access, erasure and rectification.
- Right to withdraw consent.
- Right to lodge a complaint to supervisory authority.

To facilitate the exercise of these rights, the following consent procedure will be adopted when necessary:

- A public announcement for volunteering subjects will be published on the project's web page and related social media accounts.
- Each subject will be contacted with the employed personnel of the inviting project participant at least one week before the actual session.
- If the invitation is accepted the subject will be first informed by the written information, including information about data subject rights, that will be sent to the subject via email or another communication environment as per their preference (at least one week before the session).

The HARPOCRATES project will comply with ethical standards, Horizon Europe guidelines and relevant national and EU legislation. Partners have knowledge of and compliance experience of EC ethics and legal requirements for EU-funded research projects. Trilateral has extensive experience in leading ethics and data protection efforts in H2020 and previous framework programmes research projects. The HARPOCRATES consortium will ensure respect for human dignity, fair distribution of research benefits and burden and protection of the values, rights and interests of the research participants. Major ethical principles and regulations followed in HARPOCRATES are the following:

- Horizon Europe Rules of Participation which include ethical guidelines in Article 19 (Ethics).
- The General Data Protection Regulation EU 2016/679 (GDPR) superseding the Data Protection Directive 95/46/EC.
- Directive 2002/58/EC, concerning the processing of personal data and the protection of privacy in the electronic communications sector, as modified by Directive 2009/136/EC. All national data protection and privacy laws for pilot countries will be also followed.
- The Horizon Europe Model Grant Agreement rules on ethical issues in Article 14 (Ethics) and Annex 5 (gender equality).
- the social responsibility principle, as outlined in the SATORI CEN Workshop Agreement 17145 (2017).
- Responsible Research and Innovation (RRI) principles, including diversity and inclusion, anticipation and reflection, openness and transparency, responsiveness and open access to scientific knowledge.



- Article 8 of the European Convention on Human Rights (ECHR) provides a right to respect for one's "private and family life, his home and his correspondence", subject to certain restrictions
- National legislation and other data protection related regulations applicable to HARPOCRATES partners. If necessary, the partners will request a permission from their National Data Protection Authorities.
- The results of the Ethics Review and subsequent Ethics Requirements, if any, as they will be entered into the Annex of the Grant Agreement, will be an essential part of the ethics management of the project.

## Bibliography

- [1] Michel Abdalla, Florian Bourse, Angelo De Caro, and David Pointcheval. Simple functional encryption schemes for inner products. In *IACR International Workshop on Public Key Cryptography*, pages 733–751. Springer, 2015. [5](#)
- [2] Michel Abdalla, Dario Catalano, Dario Fiore, Romain Gay, and Bogdan Ursu. Multi-input functional encryption for inner products: Function-hiding realizations and constructions without pairings. In Hovav Shacham and Alexandra Boldyreva, editors, *Advances in Cryptology – CRYPTO 2018*, Cham, 2018. Springer International Publishing.
- [3] Michel Abdalla, Romain Gay, Mariana Raykova, and Hoeteck Wee. Multi-input inner-product functional encryption from pairings. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 601–626. Springer, 2017. [5](#)
- [4] Archita Agarwal, Maurice Herlihy, Seny Kamara, and Tarik Moataz. Encrypted databases for differential privacy. *Proceedings on Privacy Enhancing Technologies*, 2019(3):170–190, 2019. [9](#)
- [5] Saikrishna Badrinarayanan, Vipul Goyal, Aayush Jain, and Amit Sahai. Verifiable functional encryption. In *International Conference on the Theory and Application of Cryptology and Information Security*. Springer, 2016. [5](#)
- [6] Alexandros Bakas and Antonis Michalas. Multi-input functional encryption: Efficient applications from symmetric primitives. In *19th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, TrustCom 2020, Guangzhou, China, December 29, 2020 - January 1, 2021*, pages 1105–1112. IEEE, 2020.
- [7] Alexandros Bakas, Antonis Michalas, and Amjad Ullah. (f)unctional sifting: A privacy-preserving reputation system through multi-input functional encryption. In *Secure IT Systems - 25th Nordic Conference, NordSec 2020, Virtual Event, November 23-24, 2020*, *Proceedings*, pages 111–126. Springer, 2020. [5](#)
- [8] Peva Blanchard, El Mahdi El Mhamdi, Rachid Guerraoui, and Julien Stainer. Machine learning with adversaries: Byzantine tolerant gradient descent. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17*, page 118–128, Red Hook, NY, USA, 2017. Curran Associates Inc. [11](#)
- [9] D. Boneh and M. Franklin. Identity-based encryption from the weil pairing. In *International cryptology conference*. Springer, 2001. [5](#)
- [10] Dan Boneh, Amit Sahai, and Brent Waters. Functional encryption: Definitions and challenges. In *Theory of Cryptography Conference*, pages 253–273. Springer, 2011. [5](#)
- [11] Zvika Brakerski. Fully homomorphic encryption without modulus switching from classical gapsvp. In *Annual Cryptology Conference*, pages 868–886. Springer, 2012. [5](#)
- [12] Zvika Brakerski, Craig Gentry, and Vinod Vaikuntanathan. (leveled) fully homomorphic encryption without bootstrapping. *ACM Transactions on Computation Theory (TOCT)*, 6(3):1–36, 2014. [5](#)
- [13] Zvika Brakerski, Ilan Komargodski, and Gil Segev. Multi-input functional encryption in the private-key setting: Stronger security from weaker assumptions. *Journal of Cryptology*, 31(2):434–520, 2018. [5](#)
- [14] Hervé Chabanne, Amaury de Wargny, Jonathan Milgram, Constance Morel, and Emmanuel Prouff. Privacy-preserving classification on deep neural network. *IACR Cryptol. ePrint Arch.*, 2017:35, 2017. [10](#)
- [15] Girish Chandrashekar and Ferat Sahin. A survey on feature selection methods. *Comput. Electr. Eng.*, 40(1):16–28, January 2014. [10](#)
- [16] L. Chen, H. Wang, Z. Charles, and D. Papailiopoulos. Draco: Byzantine-resilient distributed training via redundant gradients, 2018. [11](#)
- [17] Ilaria Chillotti, Nicolas Gama, Mariya Georgieva, and Malika Izabachène. Tffe: fast fully homomorphic encryption over the torus. *Journal of Cryptology*, 33(1):34–91, 2020. [5](#)
- [18] Edward Chou, Josh Beal, Daniel Levy, Serena Yeung, Albert Haque, and Li Fei-Fei. Faster cryptonets: Leveraging sparsity for real-world encrypted inference. *arXiv preprint arXiv:1811.09953*, 2018. [10](#)
- [19] J. DesLauriers, T. Kiss, R. Ariyattu, H.V. Dang, A. Ullah, J. Bowden, D. Krefting, G. Pierantoni, and G. Terstyanszky. Cloud apps to-go: Cloud portability with toasca and micado. *Concurrency and Computation: Practice and Experience*, 33(19):e6093, 2021. [11](#)
- [20] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006. [6](#)
- [21] Cynthia Dwork, Moni Naor, Toniann Pitassi, and Guy N Rothblum. Differential privacy under continual observation. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 715–724, 2010. [6](#)
- [22] Junfeng Fan and Frederik Vercauteren. Somewhat practical fully homomorphic encryption. *IACR Cryptol. ePrint Arch.*, 2012, 2012. [5](#)
- [23] Craig Gentry. Fully homomorphic encryption using ideal lattices. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 169–178, 2009. [5](#)
- [24] Nele Gerrits, Bart Elen, Toon Van Craenendonck, Danai Triantafyllidou, Ioannis N. Petropoulos, Rayaz A. Malik, and Patrick De Boever. Age and sex affect deep learning prediction of cardiometabolic risk factors from retinal images. *Scientific Reports*, 10(1):9432, 2020. [8](#)
- [25] Ran Gilad-Bachrach, Nathan Dowlin, Kim Laine, Kristin Lauter, Michael Naehrig, and John Wernsing. Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy. In *International Conference on Machine Learning*, 2016. [6](#), [10](#)
- [26] Shafi Goldwasser, S Dov Gordon, Vipul Goyal, Abhishek Jain, Jonathan Katz, Feng-Hao Liu, Amit Sahai, Elaine Shi, and Hong-Sheng Zhou. Multi-input functional encryption. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 578–602. Springer, 2014. [5](#), [8](#)
- [27] Shafi Goldwasser, Yael Tauman Kalai, Raluca Ada Popa, Vinod Vaikuntanathan, and Nickolai Zeldovich. How to run turing machines on encrypted data. In *Annual Cryptology Conference*, pages 536–553. Springer, 2013.
- [28] Sergey Gorbunov, Vinod Vaikuntanathan, and Hoeteck Wee. Functional encryption with bounded collusions via multi-party computation. In *Annual Cryptology Conference*, pages 162–179. Springer, 2012. [5](#)
- [29] Jane Henriksen-Bulmer and Sheridan Jeary. Re-identification attacks: a systematic literature review. *Int. J. Inf. Manag.*, 36(6), 2016. [8](#)
- [30] Ehsan Hsamifard, Hassan Takabi, and Mehdi Ghasemi. Cryptodl: towards deep learning over encrypted data. In *Annual Computer Security Applications Conference, Los Angeles, California, USA, 2016*. [10](#)

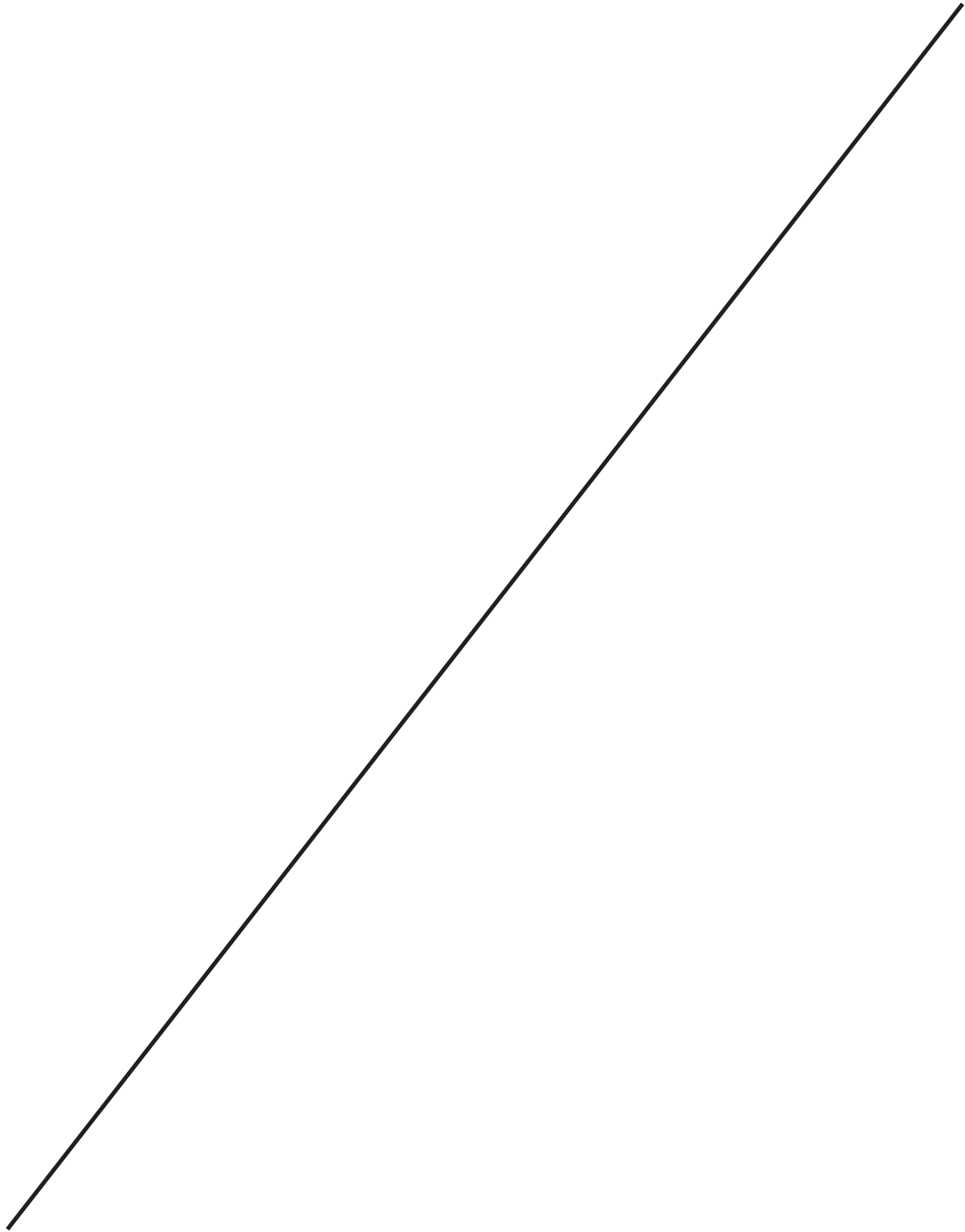


- [31] Seny Kamara, Tarik Moataz, and Olya Ohrimenko. Structured encryption and leakage suppression. In *Annual International Cryptology Conference*, pages 339–370. Springer, 2018. [9](#)
- [32] L. Lamport, R. Shostak, and M. Pease. The byzantine generals problem. *ACM Trans. Program. Lang. Syst.*, 4(3), 1982. [11](#)
- [33] Yehuda Lindell and Benny Pinkas. Privacy preserving data mining. In *Annual International Cryptology Conference*. Springer, 2000. [6](#)
- [34] P.C.T. Lipton, S.I. Moser, D.V. Palma, and T.I. Spatzier. Topology and Orchestration Specification for Cloud Applications Version 1.0. Standard, OASIS Standard, 2013. [12](#)
- [35] Luca Melis, Congzheng Song, Emiliano De Cristofaro, and Vitaly Shmatikov. Exploiting unintended feature leakage in collaborative learning. In *2019 IEEE Symposium on Security and Privacy (SP)*, pages 691–706, 2019. [11](#)
- [36] El Mhamdi, Rachid Guerraoui, and Sébastien Rouault. The hidden vulnerability of distributed learning in byzantium, 2018. [11](#)
- [37] M. Naehrig, K. Lauter, and V. Vaikuntanathan. Can homomorphic encryption be practical? In *Proceedings of the 3rd ACM workshop on Cloud computing security workshop*, 2011. [5](#), [9](#), [10](#)
- [38] Pascal Paillier. Public-key cryptosystems based on composite degree residuosity classes. In *International conference on the theory and applications of cryptographic techniques*, pages 223–238. Springer, 1999. [9](#)
- [39] Le Trieu Phong and Tran Thi Phuong. Privacy-preserving deep learning via weight transmission. *IEEE Transactions on Information Forensics and Security*, 14(11):3003–3015, 2019. [11](#)
- [40] Ronald L Rivest, Len Adleman, Michael L Dertouzos, et al. On data banks and privacy homomorphisms. *Foundations of secure computation*, 4(11):169–180, 1978. [5](#)
- [41] Edouard Dufour Sans, Romain Gay, and David Pointcheval. Reading in the dark: Classifying encrypted digits with functional encryption. *IACR Cryptology ePrint Archive*, 2018:206, 2018. [5](#)
- [42] Jinyun So, Basak Guler, and A. Salman Avestimehr. Codedprivateml: A fast and privacy-preserving framework for distributed machine learning. *IEEE Journal on Selected Areas in Information Theory*, 2(1). [6](#)
- [43] C. Sun, A. Shrivastava, S. Singh, and A. Gupta. Revisiting unreasonable effectiveness of data in deep learning era, 2017. [11](#)
- [44] Z. Tang, S. Shi, X. Chu, W. Wang, and B. Li. Communication-efficient distributed deep learning: A comprehensive survey, 2020. [11](#)
- [45] S. Truex, N. Baracaldo, A. Anwar, T. Steinke, H. Ludwig, R. Zhang, and Y. Zhou. A hybrid approach to privacy-preserving federated learning. In *Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security, AISec'19*, 2019. [11](#)
- [46] Jaideep Vaidya, Hwanjo Yu, and Xiaoqian Jiang. Privacy-preserving svm classification. *Knowledge and Information Systems*, 2008. [6](#)
- [47] P. Vepakomma, A. Singh, O. Gupta, and R. Raskar. Nopeek: Information leakage reduction to share activations in distributed deep learning, 2020. [6](#)
- [48] Praneeth Vepakomma, Otkrist Gupta, Tristan Swedish, and Ramesh Raskar. Split learning for health: Distributed deep learning without sharing raw patient data, 2018. [6](#)
- [49] Y. You, J. Li, S. Reddi, J. Hseu, S. Kumar, S. Bhojanapalli, X. Song, J. Demmel, K. Keutzer, and C-J. Hsieh. Large batch optimization for deep learning: Training bert in 76 minutes, 2020. [11](#)
- [50] Chun-Hsien Yu, Chun-Nan Chou, and Emily Chang. Distributed layer-partitioned training for privacy-preserved deep learning. In *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, 2019. [11](#)
- [51] Baobao Zhang and Allan Dafoe. Artificial intelligence: American attitudes and trends (january 2019), Jan 2019. [8](#)
- [52] Ligeng Zhu and Song Han. *Deep Leakage from Gradients*, pages 17–31. Springer International Publishing, Cham, 2020. [11](#)



---

 Associated with document Ref. Ares(2022)4525586 - 20/06/2022



fdec76f2



ANNEX 2

ESTIMATED BUDGET FOR THE ACTION

Forms of funding	Estimated eligible <sup>1</sup> costs (per budget category)										Estimated EU contribution <sup>2</sup>				
	Direct costs					Indirect costs					Total costs	Funding rate % <sup>4</sup>	Maximum EU contribution <sup>3</sup>	Requested EU contribution	Maximum grant amount <sup>6</sup>
	A. Personnel costs		B. Subcontracting costs		C. Purchase costs		D. Other cost categories		E. Indirect costs <sup>5</sup>						
A.1 Employees (or equivalent)	Unit costs (usual accounting practices)	A.2 Natural persons under direct contract	Unit costs <sup>7</sup>	B.1 Travel and subsistence	C.1 Equipment	C.2 Other goods, works and services	D.2 Internally invoiced goods and services	E. Indirect costs	Flat-rate costs <sup>8</sup>						
A.3 Seconded persons	a1	a2	a3	b	c1	c2	c3	d2	e = 0,25 * (a1 + a2 + a3 + c1 + c2 + c3)	f = a + b + c + d + e	U	g = f * U%	h	m	
1- TUNI	0,00	562.800,00	0,00	0,00	20.000,00	0,00	33.000,00	0,00	153.950,00	769.750,00	100	769.750,00	769.750,00	769.750,00	
2- RISE	392.400,00	0,00	0,00	0,00	20.000,00	0,00	9.000,00	0,00	105.350,00	526.750,00	100	526.750,00	526.750,00	526.750,00	
3- TRIE	227.500,00	0,00	0,00	0,00	20.000,00	0,00	4.000,00	0,00	62.875,00	314.375,00	100	314.375,00	314.375,00	314.375,00	
4- ZENTRIX LAB LLC	211.200,00	0,00	0,00	0,00	20.000,00	0,00	4.000,00	0,00	58.800,00	294.000,00	100	294.000,00	294.000,00	294.000,00	
5- CBIT	312.300,00	0,00	0,00	0,00	20.000,00	0,00	4.000,00	0,00	84.075,00	420.375,00	100	420.375,00	420.375,00	420.375,00	
6- CHARTE	158.400,00	0,00	0,00	0,00	9.000,00	9.000,00	3.600,00	0,00	45.000,00	225.000,00	100	225.000,00	225.000,00	225.000,00	
7- DMG	155.040,00	0,00	0,00	0,00	10.000,00	9.000,00	3.600,00	0,00	44.410,00	222.050,00	100	222.050,00	222.050,00	222.050,00	
8- SARGA	128.000,00	0,00	0,00	0,00	14.000,00	0,00	3.600,00	0,00	36.400,00	182.000,00	100	182.000,00	182.000,00	182.000,00	
9- VR-IC/dep	211.200,00	0,00	0,00	30.500,00	15.000,00	0,00	4.000,00	0,00	57.500,00	318.250,00	100	318.250,00	318.250,00	318.250,00	
10- UEF	180.000,00	0,00	0,00	0,00	10.000,00	0,00	14.000,00	0,00	51.000,00	255.000,00	100	255.000,00	255.000,00	255.000,00	
11- UP	163.200,00	0,00	0,00	0,00	10.000,00	10.000,00	4.000,00	0,00	46.800,00	234.000,00	100	234.000,00	234.000,00	234.000,00	
12- S2 GRUPPO	179.200,00	0,00	0,00	0,00	20.000,00	0,00	4.000,00	0,00	50.800,00	254.000,00	100	254.000,00	254.000,00	254.000,00	
13- 99985250															
<b>Σ consortium</b>	2.318.440,00	562.800,00	0,00	30.500,00	188.000,00	28.000,00	90.800,00	0,00	797.010,00	4.015.550,00		4.015.550,00	4.015.550,00	4.015.550,00	

<sup>1</sup> See Article 6 for the eligibility conditions. All amounts must be expressed in EUR (see Article 21 for the conversion rules).

<sup>2</sup> The consortium remains free to decide on a different internal distribution of the EU funding (via the consortium agreement, see Article 7).

<sup>3</sup> Indirect costs already covered by an operating grant (received under any EU funding programme) are negligible (see Article 6.3). Therefore, a beneficiary/affiliated entity that receives an operating grant during the action duration cannot declare indirect costs for the year(s) reporting period(s) covered by the operating grant, unless they can demonstrate that the operating grant does not cover any costs of the action. This requires specific accounting tools. Please immediately contact us via the EU Funding & Tenders Portal for details.

<sup>4</sup> See Data Sheet for the funding rate(s).

<sup>5</sup> This is the theoretical amount of the EU contribution to costs, if the reimbursement rate is applied to all the budgeted costs. This theoretical amount is then capped by the maximum grant amount.

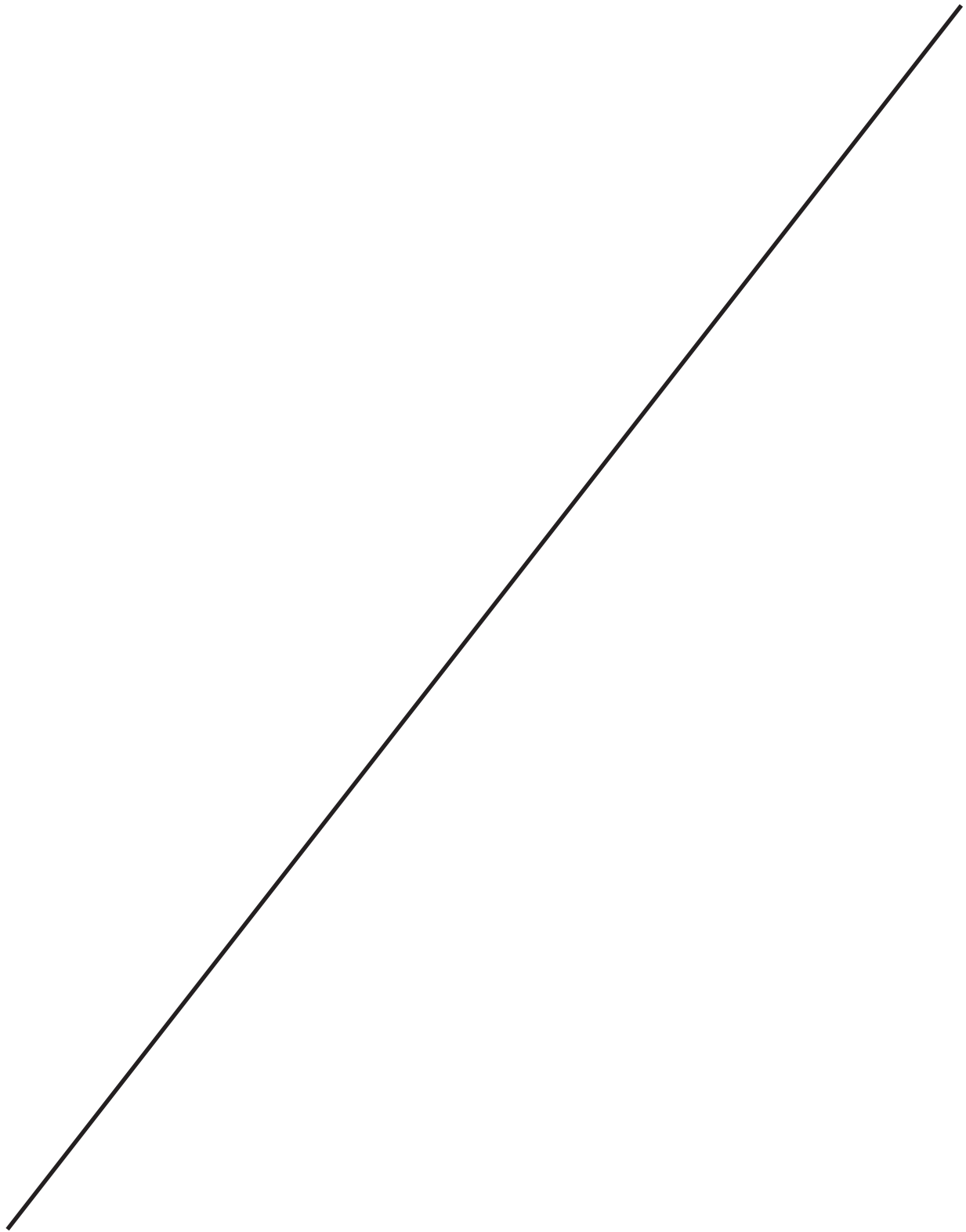
<sup>6</sup> The maximum grant amount is the maximum grant amount decided by the EU. It normally corresponds to the requested grant, but may be lower.

<sup>7</sup> See Annex 2a. Additional information on the estimated budget for the details (units, costs per unit).

<sup>8</sup> See Data Sheet for the flat-rate.







fdec76f2

