A SCALABLE AND PRACTICAL PRIVACY-PRESERVING FRAMEWORK

1st EASiNet Scientific Remote Meeting, September 14^{th}, 2024

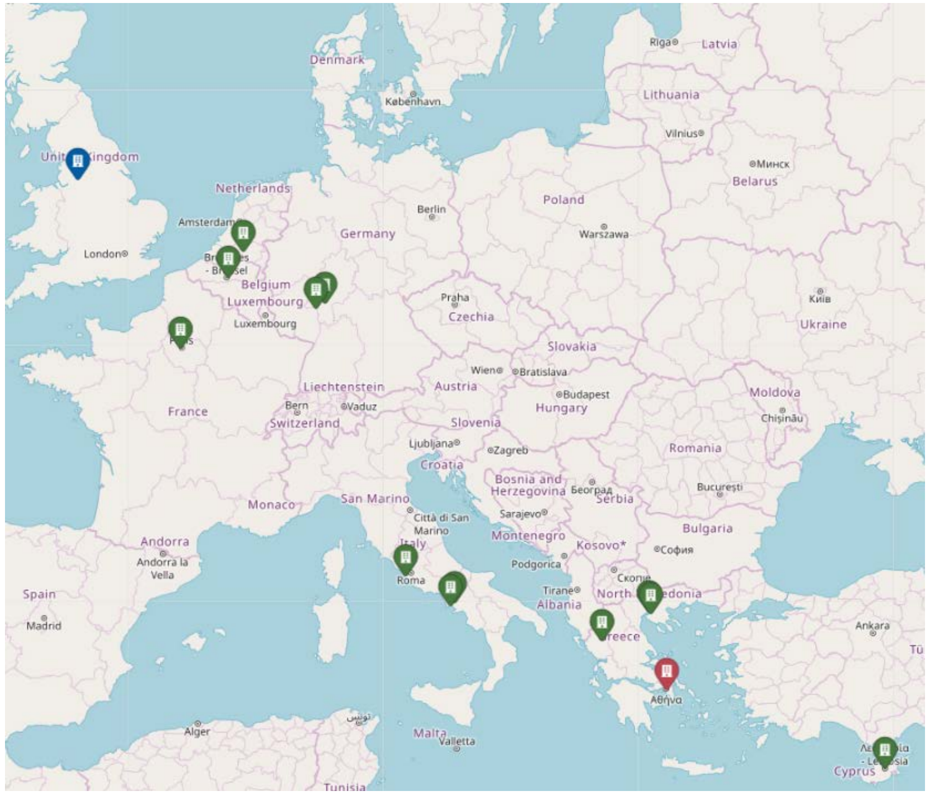
Salvatore D'Antonio

TRUSTUP-University of Naples Parthenope

encrypt

ENCRYPT Facts and Figures

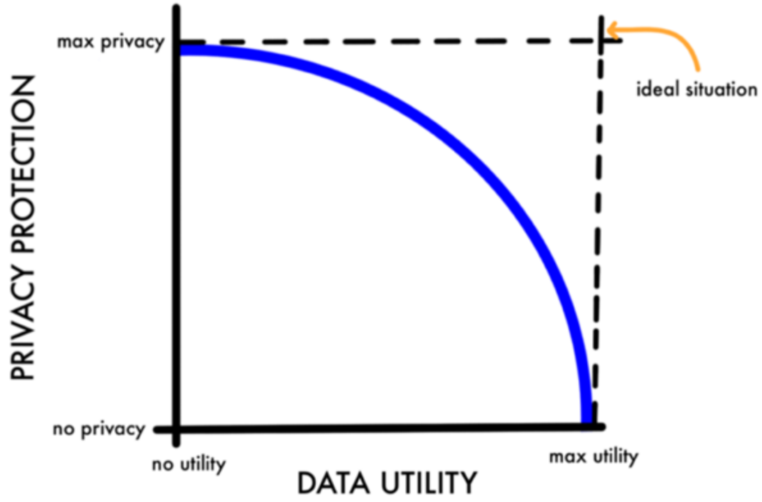
- **Project Short Name:** ENCRYPT (A SCALABLE AND PRACTICAL PRIVACY-PRESERVING FRAMEWORK)
- **Grand Agreement ID**: 101070670
- **HORIZON-CL3-2021-CS-01-04** - Scalable privacy-preserving technologies for cross-border federated computation in EU involving personal data
- **Funding Scheme:** Research and Innovation Action (RIA)
- **Total Funding:** 4,392,540 €
- **Duration:** 36 Months (July 2022 – June 2025)
- **Consortium:** 14 partners, 8 countries
  - ✓ 1 start-up (TRUSTUP)
  - ✓ 3 x SMEs (EXUS, 8BELLS, DBC)
  - ✓ 2 x Enterprises (ENG, EPIBANK)
  - ✓ 8 Research Institutes (CERTH, AUTH, UNIMAN, TIU, CEA, UNINA, GUF, UMC-Mainz)
- **Coordinator:** EXUS SOFTWARE MONOPROSOPI ETAIRIA PERIORISMENIS EVTHINIS (EXUS) – Greece
- **Website**: https://encrypt-project.eu/



encrypt

Challenges and ENCRYPT Vision

- Huge amounts of data from sectors like healthcare, banking and finance, social media
  - ✓ Researchers and service providers working with sensitive and personal data need to process them in a privacy-preserving fashion
  - ✓ State-of the-art Privacy-Preserving technologies, such as Homomorphic Encryption, Differential Privacy, suffer from scalability issues
  - ✓ Trade-off between privacy protection and data usability



- ENCRYPT will deliver a scalable, practical, adaptable privacy-preserving framework facilitating the GDPR-compliant processing of data stored in federated cross-border data spaces by exploiting
  - Privacy-preserving computation technologies
  - Supporting technologies, including a recommendations system and a methodological framework to assess the level of privacy risk and the impact of a data breach to the organization
  - Validation in 3 real-world use cases

encrypt

ENCRYPT High Level Objectives

1. To **improve the performance** of Privacy-Preserving technologies and their applicability to GDPR compliant, cross-border federated processing of sensitive data

2. To **improve the user-friendliness** of Privacy-Preserving technologies facilitating their identification, understanding, selection, and adoption **by all involved actors**

3. To foster, and inherently support **interoperability between** Privacy-Preserving **technologies operating and processing similar data types** across different organisations as well as across different sectors

4. To promote GDPR-compliant common **European Data Spaces** and facilitate the **exchange of Cyber Threat Intelligence information**, liaising with relevant initiatives and projects

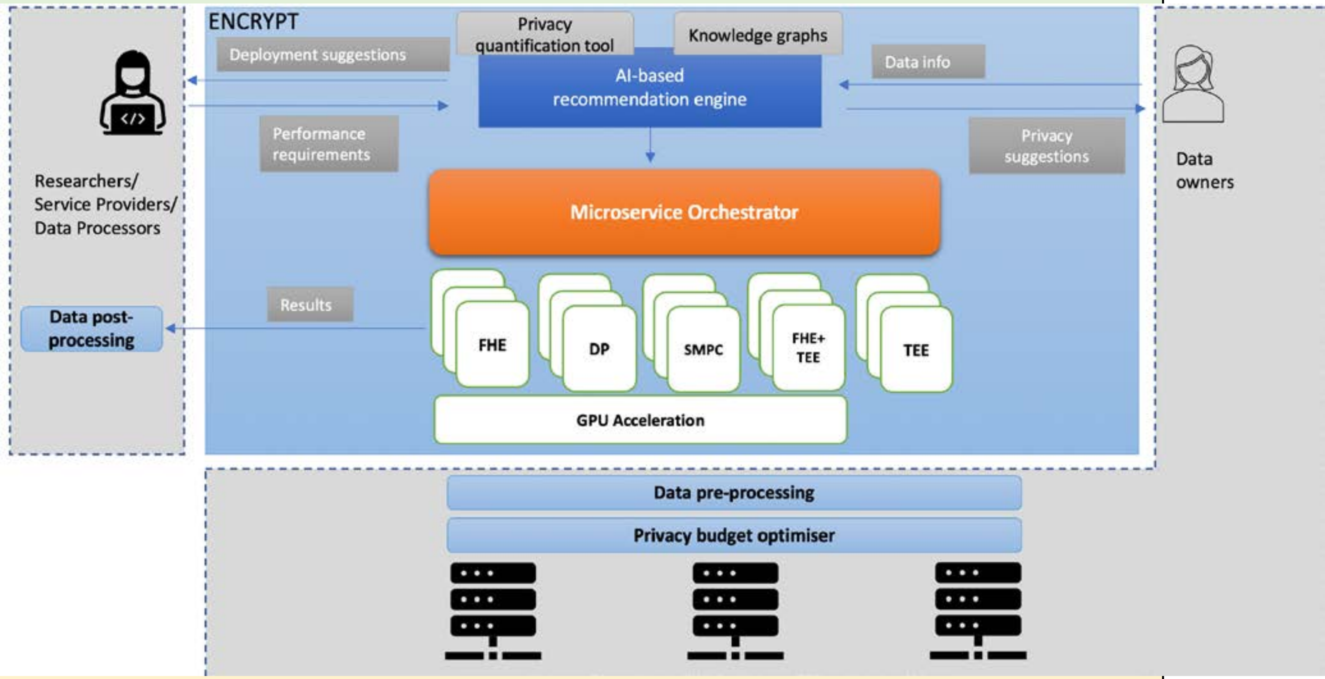
5. To ensure the **usability** of the developed solutions, by **co-designing them in cooperation with end-users**, and validating them in **realistic use cases** including federated data infrastructures

6. To **strengthen the community of** privacy-preserving technology developers by disseminating and exploiting open-source project results

ENCRYPT Key Technologies and results

Privacy –Preserving Computation Solutions

- ✓ Fully Homomorphic Encryption (FHE)
- ✓ Trusted Execution Environment (TEE)
- ✓ Differential Privacy (DP)
- ✓ Combined FHE+TEE, FHE+DP
- ✓ Acceleration Service based on GPU hardware

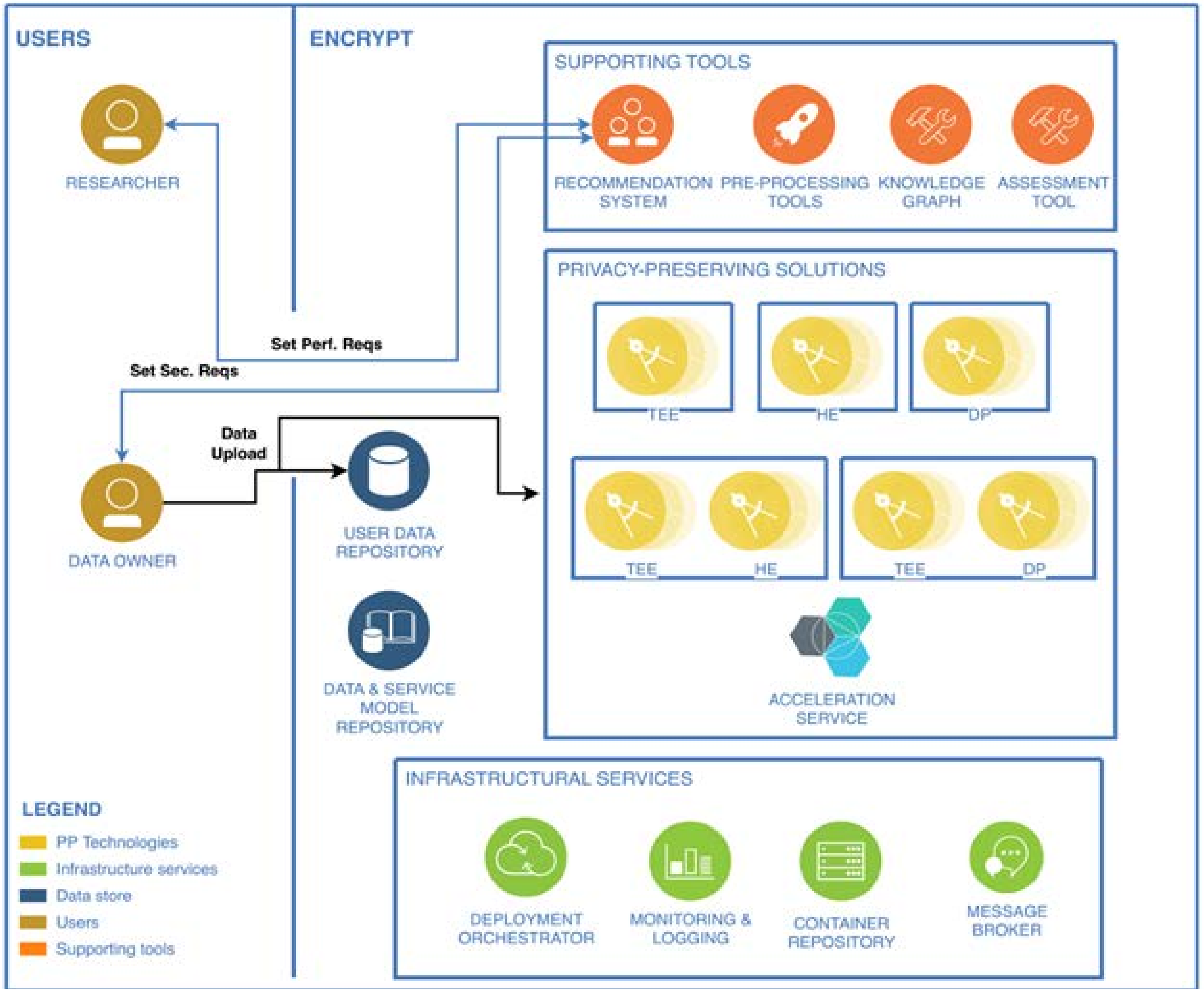


Privacy-Supporting Tools

- ✓ Advanced data-preprocessing
- ✓ Knowledge Graphs
- ✓ Methodological Framework for Privacy Risk Assessment
- ✓ AI–powered Recommendation System
- ✓ Front-end and back-end services

encrypt

ENCRYPT platform architecture - 1



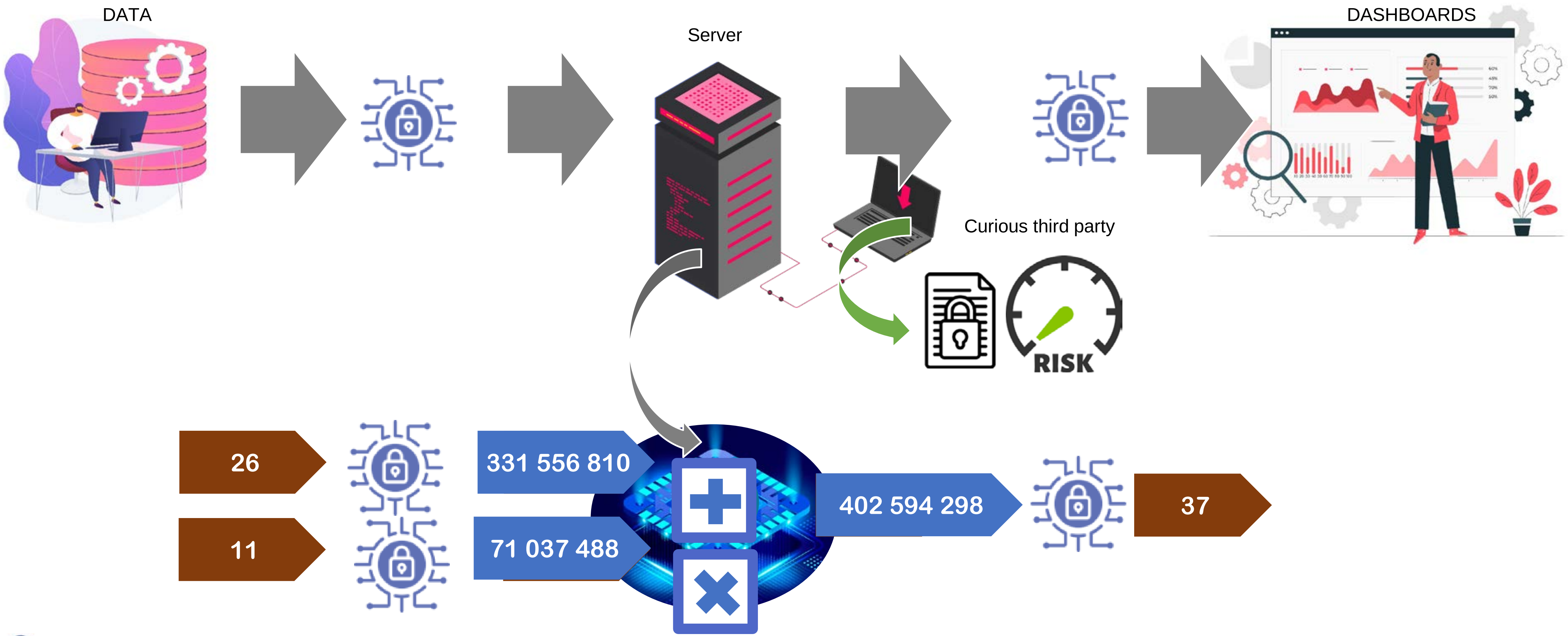
ENCRYPT platform architecture - 2

- Adoption of the RM-ODP (Reference Model of Open Distributed Processing) architectural standard
  - ✓ The Enterprise viewpoint includes a definition of ENCRYPT objectives, services, and users
  - ✓ The Computation viewpoint overviews the different components and their organization
  - ✓ The Information viewpoint defines different service information flows at all interfaces within the ENCRYPT architecture
  - ✓ The Engineering viewpoint describes the interaction across the entire distributed ENCRYPT architecture
  - ✓ The Technology viewpoint includes the identification of ENCRYPT technologies, used for the implementation, e.g., hardware and software platforms, networks, storage devices
- Use of containerization to guarantee isolation and of an event streaming platform like Kafka to ensure scalability and a loosely-coupled integration

ENCRYPT Privacy-preserving Technologies – Trusted Execution Environment

- While standard secure protocols and technologies already exist to protect data in transit, ENCRYPT is focusing on security of "data in use" and confidential computing
  - ✓ Use of Trusted Execution Environment Technologies (Intel SGX, ARM TrustZone, AMD SEV)
  - ✓ wrt Intel SGX, an extension to CPU ISA allows user-level code to allocate private regions of memory, called Secure Enclaves
  - ✓ The trust model foresees that everything outside the secure enclave is untrusted, including the OS, Hypervisor, and firmware
    - This implies that system calls are not allowed in the enclave
  - ✓ Limitation: enclave page cache (EPC) size
- Code and data are protected even from cyberattacks launched by high-privilege users

ENCRYPT Privacy-preserving Technologies – Fully Homomorphic Encryption



encrypt

ENCRYPT Privacy-preserving Technologies – Differential Privacy

- Differential Privacy (DP) ensures that the outcome of any statistical analysis does not allow to infer whether sensitive data are included in the dataset or not.
- This is achieved by altering data (or the results of queries made on a dataset) with a controlled amount of random noise.
- The addition of noise is one of the most important aspect of DP, as it obfuscates the contribution of individual data points while maintaining data utility.
- DP can be implemented using one of two main approaches - local differential privacy (LDP) and global differential privacy (GDP).
  - ✓ LDP ensures the privacy of individual's data at the source, before any data collection or analysis takes place. Noise is therefore added to user data locally on their device, and only noisy data are sent to a central server. LDP is particularly useful when users do not fully trust a data collector.
  - ✓ GDP on the other hand, applies the privacy mechanism on a centralized database/server side. In this model, a dataset is collected, stored and processed by a trusted entity and noise is added to the outputs of queries made on the database. This approach relies on the trustworthiness of the data collector.

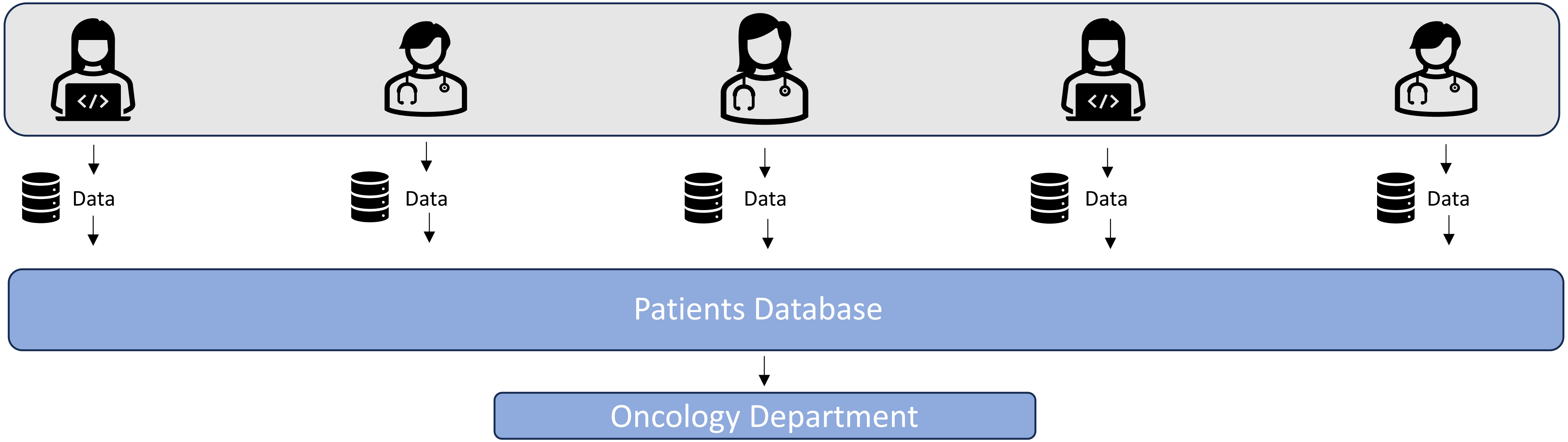
Use cases and cyber risk differentiation

Health domain: Cooperative Oncology	Cybersecurity domain: Cyber Threat Intelligence information sharing	Fintech domain: Data Analytics
Cybersecurity risk: security-induced safety implications	Cybersecurity risk: data breaches and illegal use of CTI information	Cybersecurity risk: data subject reidentification

ENCRYPT Use Case 1 – Health Domain

Cooperative oncology

- Focus on data confidentiality in the thyroid cancer diagnosis and follow-up process
- Management of patients implies the involvement of several profiles and skills, given that different types of information and data need to be collected for the same patient, both during diagnostic and follow-up phase
- Clinicians from various specialties, surgeons, biologists, technicians and even administrative staff are involved in the overall process
- In this scenario a database is set up for the storage of data and sensitive information.
  - ✓ The database with information and data of patients affected by thyroid cancer is managed by the Oncology department that is responsible for the patient's follow-up



ENCRYPT Use Case 1 – Health Domain

In case the columns exhibit specific values (i.e. indicating the presence of a certain gene mutation or metastatic disease), then the patient ID must be reported and highlighted to the Oncologist who will decide whether the patient needs an accurate and deeper follow-up.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	Patient ID	Age	Sex	Surgery	Date_surgery	Surgery_type	PTC	Hysto	RET_gene	Ras_gene	BRAF_gene	TNM_staging	T4	T	Multifocal	N	M	Tg35	Tg23	N_I131	Date_TherapyI131	TSHPre	Tgpre
2	565		1.00	1.00	03-Feb-97	1.00	1.00	1	1.00	0.00	0.00	pT3bNxMX	0.00	3.00	0.00	0.00	0.00	0.00	0.00	1.00	16-Apr-97	100.00	4.40
3	8	78.00	1.00	2.00	10-Feb-97	4.00	0.00	3	0.00	0.00	0.00	pT4aNxM1	1.00	4.00	0.00	0.00	0.00	1.00	1.00	4.00	29-Apr-97	24.00	500.00
4	302		1.00	3.00	14-Dec-94	6.00	1.00	1	0.00	0.00	0.00	pT4N1M1	1.00	4.00	0.00	1.00	1.00	1.00	1.00	2.00	14-Nov-97	45.20	111.20
5	816	78.00	1.00	2.00	15-Oct-05	5.00	1.00	1	0.00	0.00	0.00	pT4aN1bM1	1.00	4.00	0.00	1.00	0.00	1.00	1.00	1.00	05-Dec-05	36.00	964.20
6	788	70.00	1.00	2.00	09-Nov-05	4.00	0.00	3	0.00	0.00	0.00	pT2(m)NxM1	0.00	2.00	1.00	0.00	0.00	1.00	1.00	3.00	14-Dec-05	24.00	2662.00
7	600	52.00	1.00	2.00	17-Jun-03	6.00	0.00	3	0.00	0.00	0.00	pT3bNxM1	0.00	3.00	0.00	0.00	0.00	1.00	1.00	5.00	14-Jul-03	40.40	14045.00
8	219	48.00	1.00	1.00	24-Jul-96	5.00	1.00	1	0.00	0.00	0.00	pT4aN1bM1	1.00	4.00	0.00	1.00	0.00	1.00	1.00	1.00	25-Oct-96	74.30	495.00
9	451	65.00	1.00	1.00	17-Dec-03	1.00	0.00	2	0.00	0.00	0.00	pT2NxM1	0.00	2.00	0.00	0.00	0.00	1.00	1.00	3.00	26-Apr-04	24.00	1000.00
10	461	61.00	1.00	1.00	26-Jun-03	1.00	0.00	2	0.00	0.00	0.00	pT1b(m)NxM1	0.00	1.00	1.00	0.00	0.00	1.00	1.00	5.00	09-Dec-03	73.30	3000.00
11	701		0.00	1.00	24-Jun-03	1.00	0.00	3	0.00	0.00	1.00	pT4aNxM1	1.00	4.00	0.00	0.00	1.00	0.00	0.00	1.00	05-Nov-03	51.10	11.60
12	554	72.00	1.00	1.00	20-Jan-06	1.00	0.00	3	0.00	0.00	0.00	pT2NxM1	0.00	2.00	0.00	0.00	0.00	1.00	1.00	1.00	17-May-06	24.00	6424.00
13	7	82.00	1.00	1.00	17-Dec-99	5.00	0.00	2	0.00	0.00	0.00	pT4a(m)N1bM1	1.00	4.00	1.00	1.00	0.00	1.00	1.00	1.00	03-Feb-00	53.10	300.00
14	753	55.00	1.00	1.00	18-Apr-05	6.00	0.00	2	0.00	0.00	0.00	pT3bN1bM1	0.00	3.00	0.00	1.00	0.00	1.00	1.00	5.00	23-May-05	67.00	12000.00
15	420	59.00	0.00	1.00	12-Jan-09	5.00	0.00	2	0.00	0.00	0.00	PT4N0M1	0.00	1.00	0.00	0.00	0.00	1.00	1.00	3.00	11-Feb-09	80.10	5000.00
16	152	59.00	0.00	3.00	20-Jul-06	3.00	1.00	1	0.00	0.00	0.00	pT1(m)NxM1	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	25-Sep-06	24.00	0.10
17	278	78.00	0.00	1.00	21-Dec-09	1.00	0.00	2	0.00	0.00	0.00	pT4N1M1	1.00	4.00	0.00	1.00	0.00	1.00	1.00	1.00	03-Feb-10	52.00	9851.00
18	215	67.00	1.00	1.00	11-Mar-96	6.00	0.00	2	0.00	0.00	0.00	pT4aN1M1	1.00	4.00	0.00	1.00	0.00	1.00	1.00	5.00	27-May-96	24.00	175.00
19	188	42.00	0.00	1.00	17-Nov-99	2.00	0.00	3	0.00	0.00	0.00	pT2NxMx	0.00	2.00	0.00	0.00	0.00	0.00	0.00	1.00	17-Jan-01	98.00	7.20
20	61		0.00	1.00	09-Mar-98	1.00	0.00	3	0.00	0.00	0.00	T2NxM0	0.00	2.00	0.00	0.00	1.00	0.00	0.00	1.00	06-May-98	44.00	15.00
21	27	52.00	1.00	1.00	01-Jan-96	1.00	1.00	1	0.00	0.00	0.00	pT2NxMx	0.00	2.00	0.00	0.00	0.00	0.00	0.00	1.00	24-Sep-01	61.00	0.20
22	787	46.00	0.00	1.00	16-Jan-06	1.00	0.00	3	0.00	0.00	0.00	pT3Nx	0.00	3.00	0.00	0.00	0.00	1.00	1.00	3.00	14-Feb-06	44.50	6320.00
23	377	43.00	0.00	1.00	08-May-97	1.00	0.00	3	0.00	0.00	0.00	pT3N0Mx	0.00	3.00	0.00	0.00	0.00	1.00	1.00	2.00	07-Jul-97	75.00	450.00
24	355	81.00	1.00	3.00	01-Jan-96	6.00	0.00	3	0.00	0.00	0.00	T3N1b	1.00	4.00	0.00	1.00	0.00	1.00	1.00	1.00	24-Mar-97	66.00	450.00
25	423	75.00	0.00	1.00	02-Jul-03	1.00	1.00	1	0.00	0.00	0.00	pT1bNxMX	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	05-Nov-03	30.00	20.30
26	294	26.00	0.00	1.00	21-Apr-98	3.00	0.00	2	0.00	0.00	0.00	pT2NxMx	0.00	2.00	0.00	0.00	0.00	0.00	0.00	2.00	15-Jan-99	94.00	1.90
27	807	23.00	0.00	1.00	07-Dec-98	1.00	1.00	1	0.00	0.00	0.00	pT2NxMx	0.00	2.00	0.00	0.00	0.00	0.00	0.00	1.00	15-Jan-99	120.70	0.10

Fully Homomorphic Encryption for the health use-case

Main idea: *Blind checking in a table if some metadata from a patient take specific values (If at least one match is found, the patient must be reported and highlighted).*

Patient id	Sex	Age range	Surgerie s	…	…
Patient 1	1	3	0	…	…
Patient 2	0	2	2	…	…
…	…	…	…	…	…
…	…	…	…	…	…

Workflow:

- The table is filled by different medical actors from different departments.
- The table is fully homomorphically encrypted and sent to the oncology department.
- The technician or the administrative staff evaluates the encrypted data in the table by exploiting FHE-based correlation and analysis.
- If some columns show specific values (i.e. IF column R=1 or column X=1), then the patient ID must be passed to the oncologist.
- The oncologist decrypts the file and decides whether the patient needs a more accurate follow-up.

encrypt

ENCRYPT Use Case 2 – Fintech Domain

Privacy-preserving technologies are seen as the tools for meeting two requirements posed by bank institutions

- **Need 1:** Need of the financial institutions to be sure about the **security and privacy levels** required to **ensure the anonymity of their clients not only internally** to their organization, but also when they **share their data to potential 3rd parties** to perform data analysis and/or to deliver tailored software solutions for the bank's activities.
  - ✓ **Pseudonymization is not enough**.
  - ✓ **Advanced Privacy-Preserving** techniques are required.

- **Need 2:** 3rd party/entity receives these data from the financial institutions and perform AI-driven data analysis in order to deliver tailored software solutions serving the strategies and policies of the bank in specific business portfolios.
  - ✓ **Efficiency and accuracy of the AI models during training** has to be ensured when Privacy-Preserving techniques are applied

encrypt

ENCRYPT Use Case 3 – Cybersecurity Domain

- Cyber Threat Intelligence (CTI) extracts knowledge about cyberattacks
- Sharing CTI allows for enhancing cyber situational awareness and defense strategies
- Organizations are often hesitant or reluctant to share information due to concerns about sensitive data
- Implementing effective Privacy-Preserving data processing technologies can help data owners to share and exchange CTI information
- Connection with MISP platform and correlation of CTI collected information with Indicators of Compromise of targeted attacks, vulnerability information and even terrorism information stored in the MISP platform using privacy-preserving data analytics

ENCRYPT Contribution To Call Impacts

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

More user-friendly solutions for privacy-preserving processing of federated personal data registries by researchers

IMPACT

Improving privacy-preserving technologies for cyber threat intelligence and data sharing solutions

Contribution to promotion of GDPR compliant European data spaces for digital services and research

Strengthened European open-source ecosystem of developers and researchers of privacy-preserving solutions

Thanks a lot for your attention !

Contact info: salvatore.dantonio@uniparthenope.it

encrypt