HELLENIC DATA PROTECTION AUTHORITY

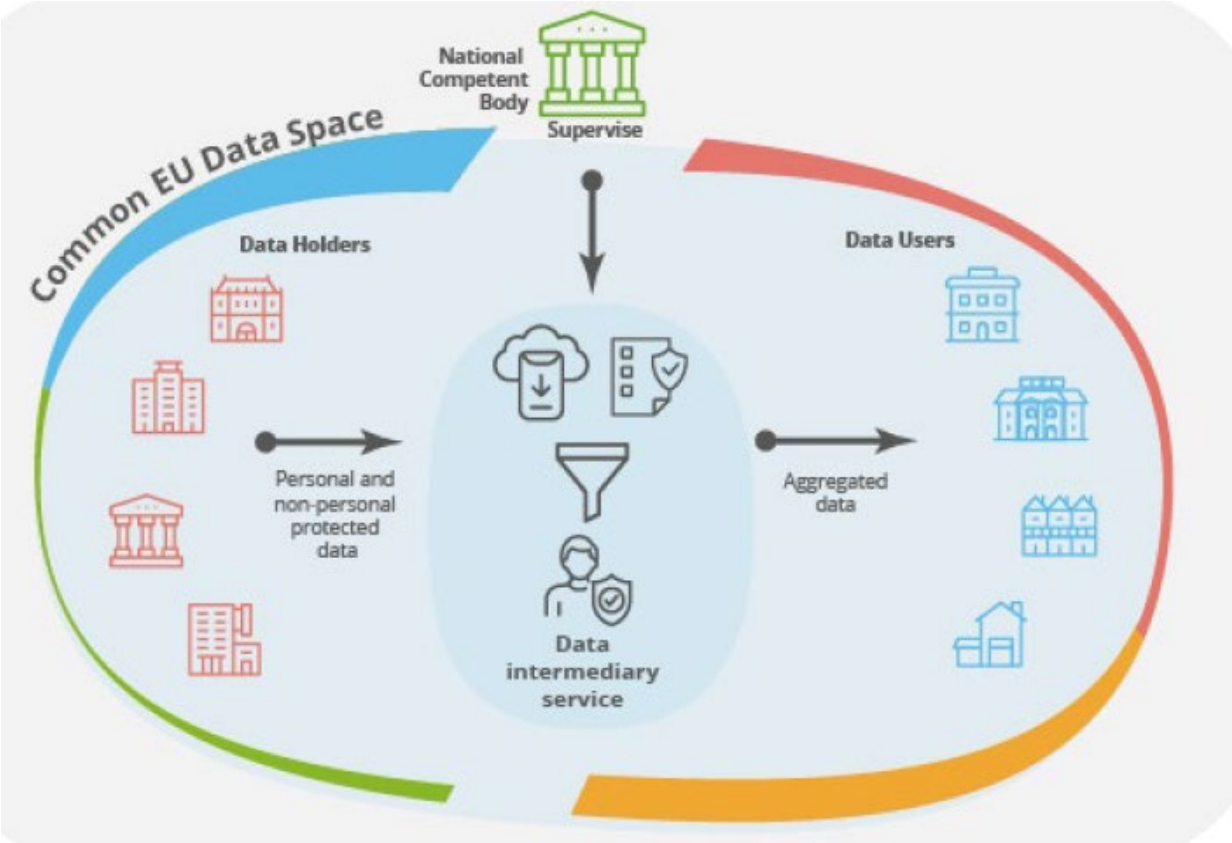
Advanced cryptographic solutions for implementing secure environments

Konstantinos Limniotis
ICT Specialist, Head of Research and Studies, HDPA
Adjunct Faculty Member – University of Athens & Open University of Cyprus

Overview

- The concept of data sharing
  - EHDS, DGA, Data Act…
- Secure processing environments
  - When?
  - How?
- The role of (advanced) cryptographic techniques
- Conclusions – Future Steps

Data sharing through intermediaries



Source: ENISA, Engineering Data Spaces, 2024

- Data holders provide aggregated/pseudonymous/anonymous data (i.e., not allowing identification of the individuals) to data users, probably through a data intermediation service
  - For well-determined (secondary) purposes (see EHDS, DGA, Data Act), in compliance with the GDPR
- ➡ Data protection engineering is essential
- The role of the intermediary service is crucial
  - Not identical roles amongst the various Acts
- The EHDS introduces the important (and quite different) notion of the Health Data Access Body

Secondary purposes in EHDS (Chapter IV – ar. 34)

- Public interest in the area of public and occupational health (…)
- Policy making and regulatory activities in the health/care sector
- Statistics related to health/care sector
- Education or teaching activities in health or care sectors
- Scientific research related to health or care sectors (…)
- Improving delivery of care, treatment optimization and providing healthcare, based on the *electronic* health data of other natural persons.

They do not need personalized information (i.e., de-identification is prerequisite)

Data intermediation service in the DGA

- **Ar. 2(10) DGA**: *Service which aims to establish commercial relationships for the purposes of data sharing between an undetermined number of data subjects and data holders on the one hand and data users on the other, through technical, legal or other means, including for the purpose of exercising the rights of data subjects in relation to personal data (…)*
- **Ar. 12 DGA:**
  - *(…) They shall not use the data for which it provides data intermediation services for purposes other than to put them at the disposal of data users (…)*
  - *(…) They may include offering additional specific tools and services to data holders or data subjects for the specific purpose of facilitating the exchange of data, such as temporary storage, curation, conversion, anonymisation and pseudonymisation, such tools being used only at the explicit request or approval of the data holder or data subject and third-party tools offered in that context not being used for other purposes; (…)*

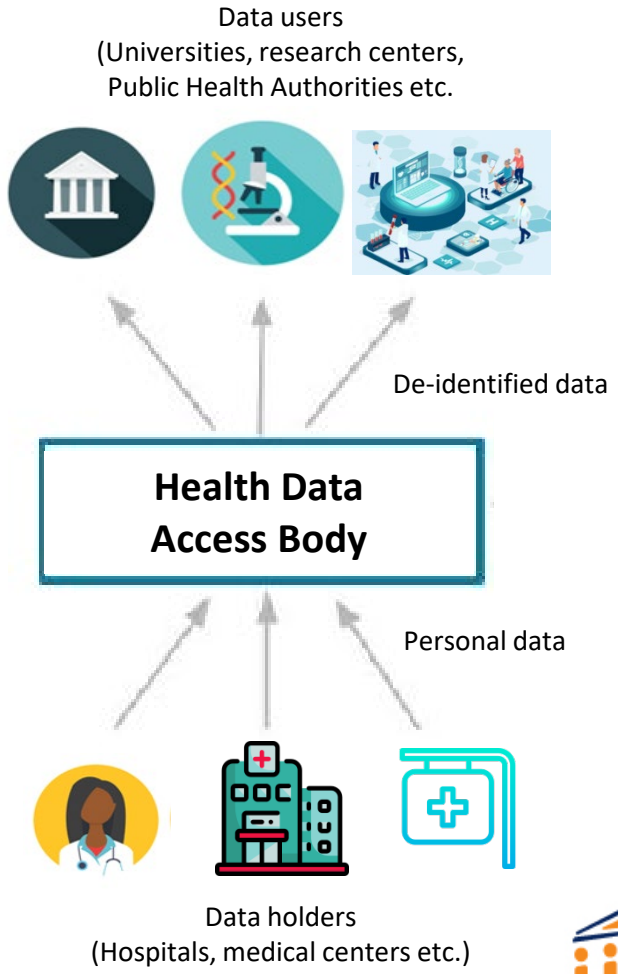
Data intermediation services in the EHDS

- EHDS: Special reference for the case of secondary use of health data
  - Ar. 32[a]: «*Member States may, by virtue of national legislation, provide that the duties of certain categories of data holders shall be fulfilled by health data intermediation entities (….)*»
  - Recital 40: *«(…) Such health data intermediation entities should be legal persons able to process and make available for secondary use electronic health data provided by data holders. Such health data intermediation entities perform different tasks than data intermediation services referred to in the DGA.»*
- However, not much are being said on what exactly health data intermediation services should do
  - Maybe (?) because:
    - They are not mandatory
    - The EHDS mandates the Health Data Access Body

Health Data Access Body (ar. 36 and 37 EHDS)

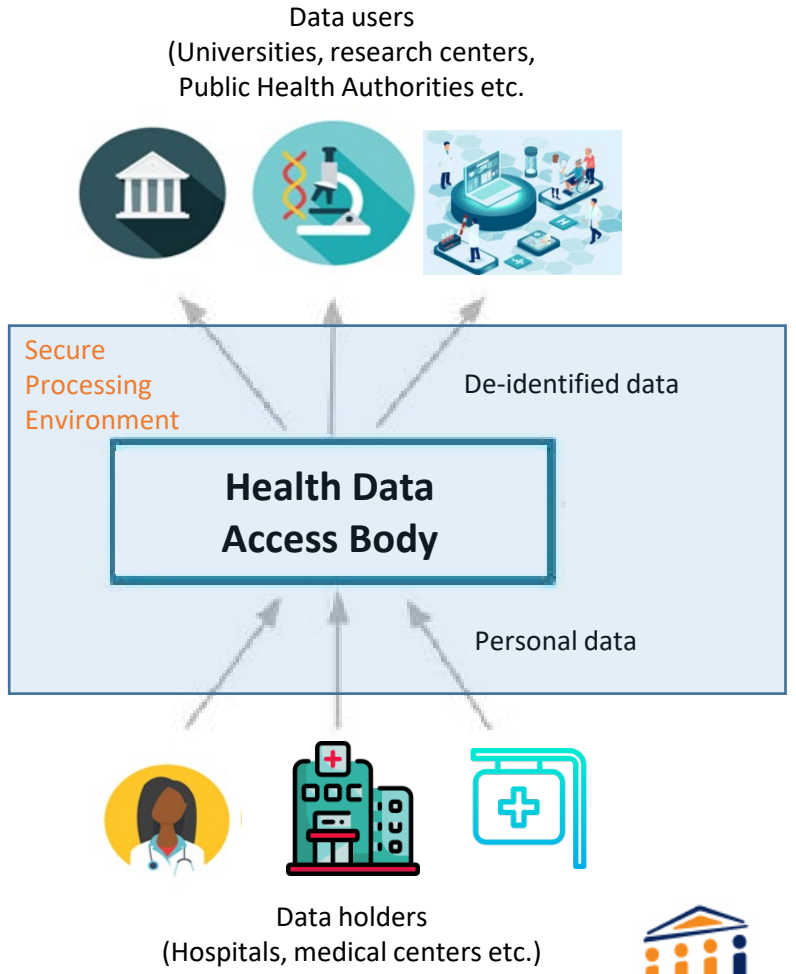
- Designated by each Member State (one or more Health Data Access Bodies)
- Decide on data access applications by data users, authorizing and issuing (if they decide so) a data permit (pursuant to the Art. 46)
- Request electronic health data from relevant health data holders pursuant to a data permit or a data request granted
- Process electronic health data such as the receiving, combination, preparation and compiling of necessary requested data from health data holders, the pseudonymisation or anonymisation of the data
- Make publicly available and easily searchable through electronic means and accessible for natural persons the conditions under which electronic health data is made available for secondary use
- Monitor and supervise compliance with the EHDS requirements by health data users and health data holders
- (….)

- So, the Health Data Access Body has a primary role
- It is under the supervision of Data Protection Authorities



The notion of the Secure Processing Environment (SPE)

- Ar. 50 EHDS: The health data access bodies shall provide access to electronic health data pursuant to a data permit only through a secure processing environment, with technical and organisational measures and security and interoperability requirements
- The SPE is defined in the DGA:
  - *"The physical or virtual environment and organisational means to ensure compliance with Union law, such as Regulation (EU) 2016/679, in particular with regard to data subjects' rights (…) and to allow the entity providing the secure processing environment to determine and supervise all data processing actions, including the display, storage, download and export of data and the calculation of derivative data through computational algorithms"*
  - Interestingly enough, the DGA does not explicitly impose an obligation on intermediation service providers to implement SPE
    - It is being mentioned (only) as an option….



Properties of a SPE (Recital 54, EHDS)

- The processing of personal data in such a secure environment should comply with Regulation (EU) 2016/679, including, where the secure environment is managed by a third party, the requirements of Article 28 and, where applicable, Chapter V.
  - Hence, data processors may be used for implementing SPE….
- Such secure processing environment should reduce the privacy risks related to such processing activities and prevent the electronic health data from being transmitted directly to the data users.
- Only non-personal electronic health data which do not contain any electronic health data should be extracted by the data users from such secure processing environment.
- (Some) more concrete provisions are given in the Art. 50 of the EHDS…

How to implement a SPE?

- It does not (seem to) exist a commonly agreed way to implement SPE
  - See, e.g., the report from the Workshop on Secure Processing Environments, 19-20 June 2023 (available online: DOI: 10.5281/zenodo.8341642)
  - Emphasis on how data users access data remotely in a secure way (e.g. through VPN)
- **However:**

  Compliance with the GDPR => Fulfilment of the data minimisation principle

  - Data minimization should be present in all phases of the process
  - For example, what if the intermediation service or the Health Data Access Body can fully support the desired tasks without having access to the original health data?
    - If this is the case, then the (secure) processing environment needs to be implemented accordingly

The role of (advanced) cryptography

- The notion of a secure processing environment is by default related with the notion of «secure computations»
- Cryptography provides the means for implementing secure computations
  - Their properties are fully in line with the desired properties of a SPE
    - Especially if a data processor is being employed…
- Some examples to be discussed (not an exhaustive list):
  - Proxy re-encryption
  - Homomorphic encryption
  - Polymorphic encryption and pseudonymisation
  - Oblivious (“Chameleon”) pseudonymisation

The role of (advanced) cryptography

- The notion of a secure processing environment is by default related with the notion of «secure computations»
- Cryptography provides the means for implementing secure computations
  - Their properties are fully in line with the desired properties of a SPE
    - Especially if a data processor is being employed…
- Some examples to be discussed (not an exhaustive list):
  - Proxy re-encryption
  - Homomorphic encryption
  - Polymorphic encryption and pseudonymisation
  - Oblivious ("Chameleon") pseudonymisation

To be considered for **any intermediation service**, regardless it is a SPE or not

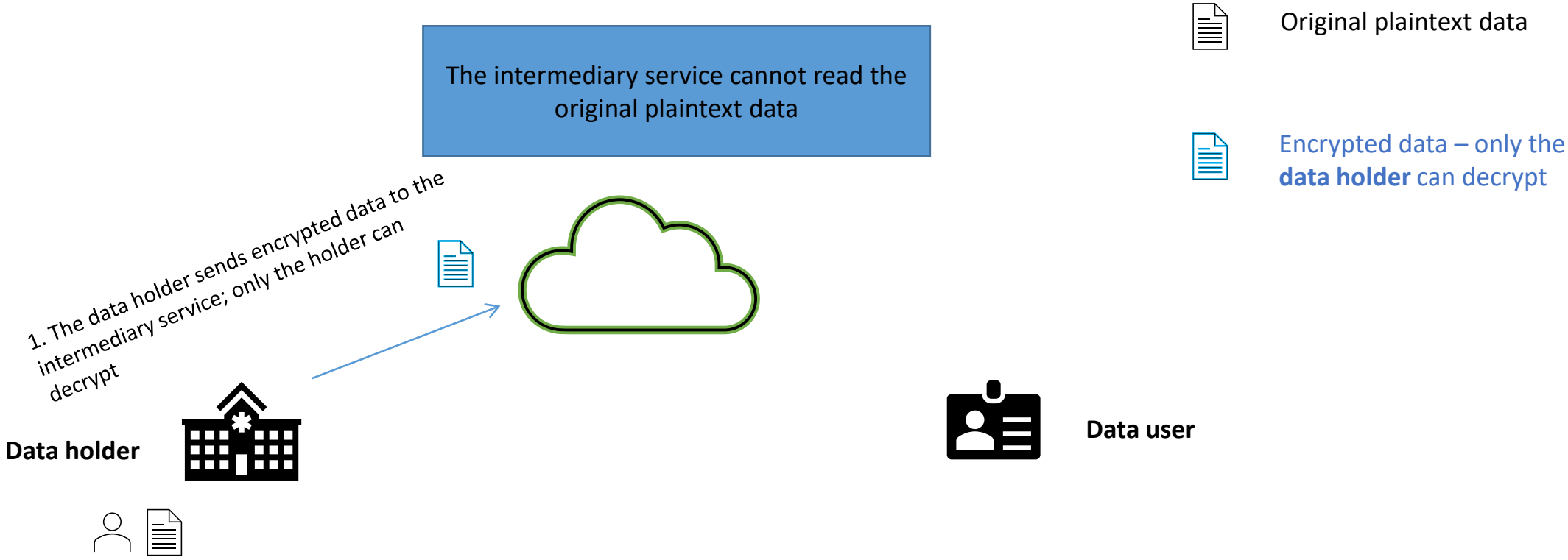
The role of (advanced) cryptography

- The notion of a secure processing environment is by default related with the notion of «secure computations»
- Cryptography provides the means for implementing secure computations
  - Their properties are fully in line with the desired properties of a SPE
    - Especially if a data processor is being employed…
- Some examples to be discussed (not an exhaustive list):
  - Proxy re-encryption
  - Homomorphic encryption
  - Polymorphic encryption and pseudonymisation
  - Oblivious ("Chameleon") pseudonymisation

For the special case of a SPE, it is **probable** that a risk-based approach will yield that (at least) <u>one of them **is necessary**</u>

- Depending on the application/context…

Proxy re-encryption

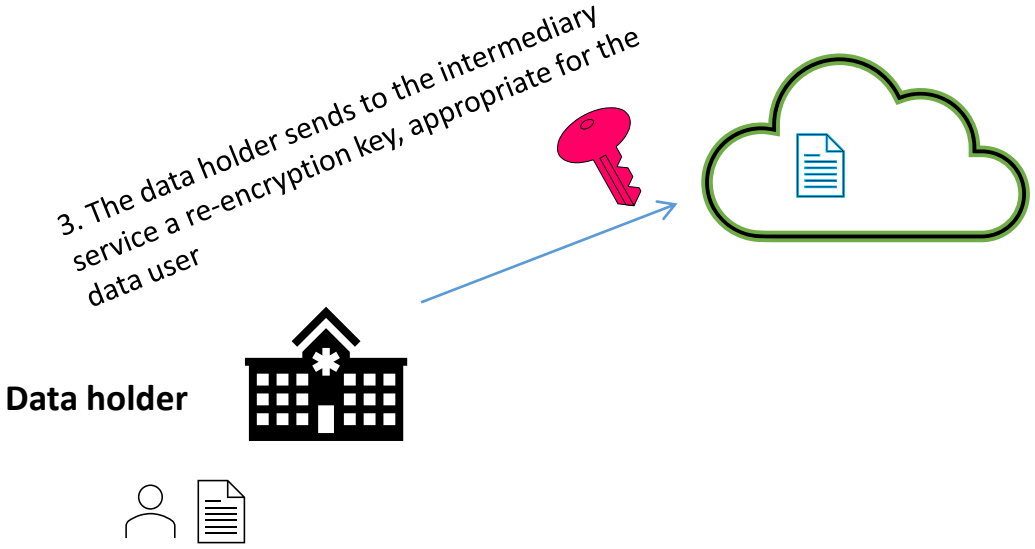


Proxy re-encryption

Original plaintext data

Encrypted data – only the **data holder** can decrypt

Proxy re-encryption



Original plaintext data

Encrypted data – only the **data holder** can decrypt

Data user

Proxy re-encryption

4. The intermediary service re-encrypts the ciphertext, generating another ciphertext that only the data user can decrypt, getting the original plaintext data.

Data holder

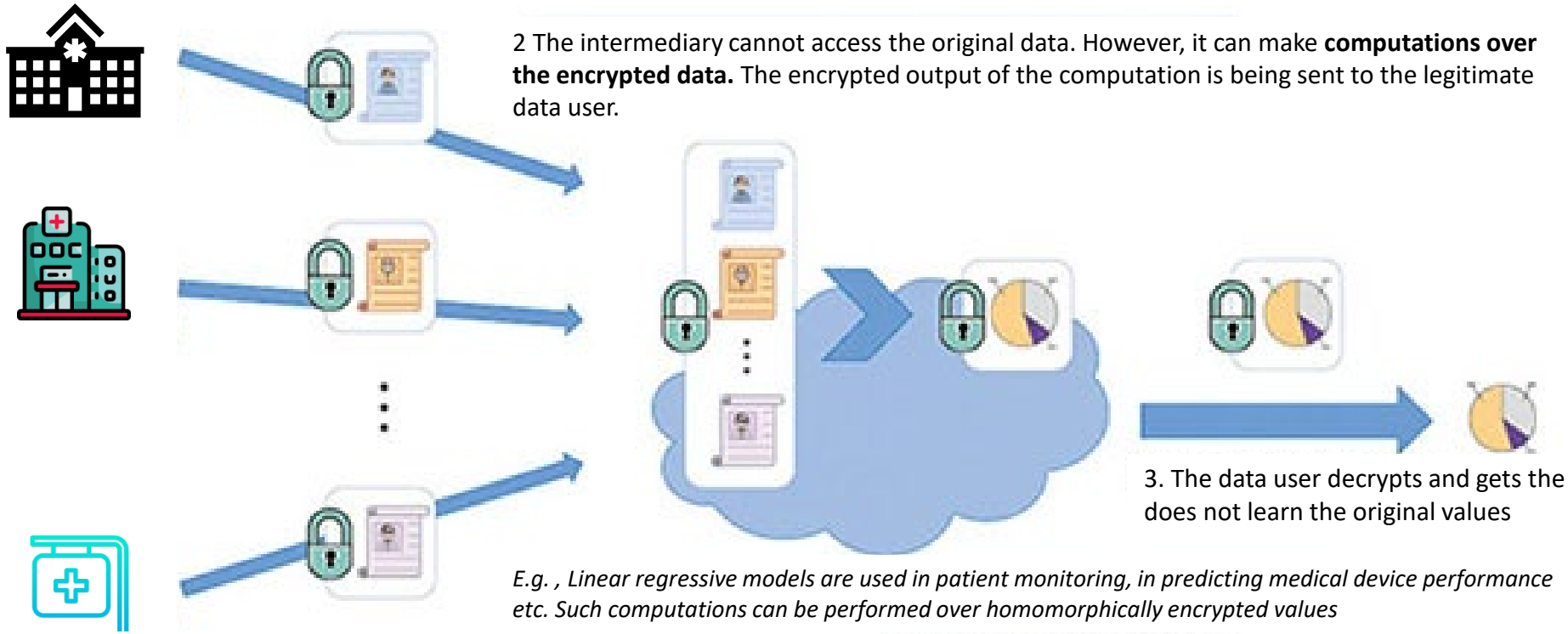
Data user

Original plaintext data

Encrypted data – only the **data holder** can decrypt

Encrypted data – only the **data user** can decrypt

Homomorphic encryption



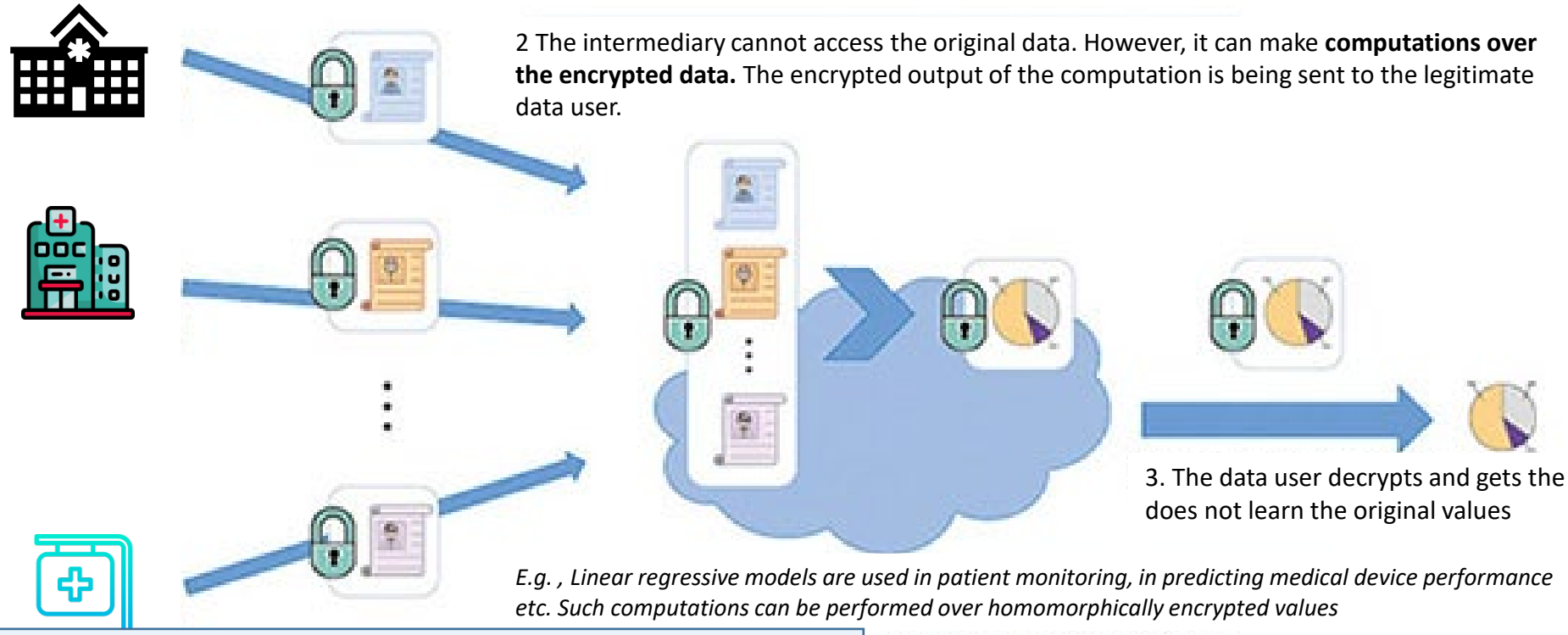
2 The intermediary cannot access the original data. However, it can make **computations over the encrypted data.** The encrypted output of the computation is being sent to the legitimate data user.

3. The data user decrypts and gets the output of the computation. It does not learn the original values

E.g. , Linear regressive models are used in patient monitoring, in predicting medical device performance etc. Such computations can be performed over homomorphically encrypted values

1. The data holders send encrypted data to the intermediary service; only the data user can decrypt.

Homomorphic encryption



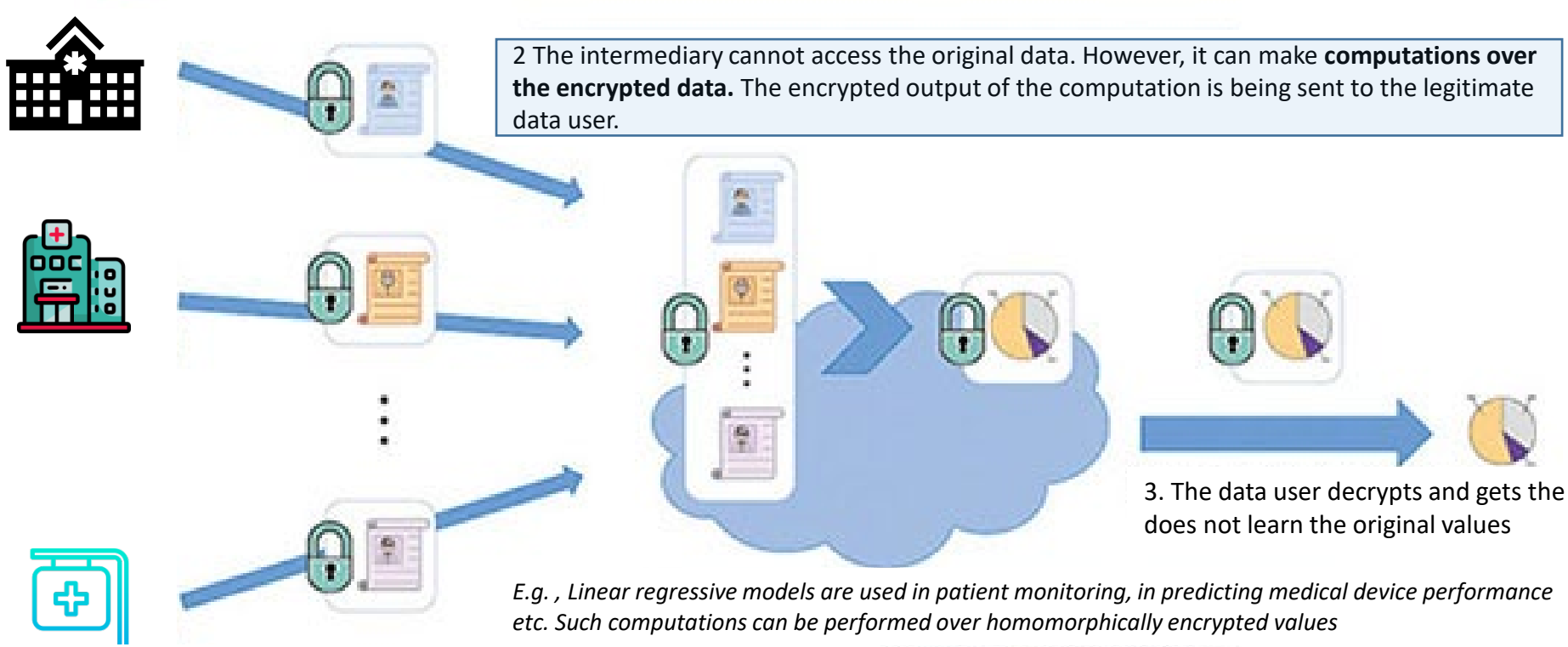
2 The intermediary cannot access the original data. However, it can make **computations over the encrypted data.** The encrypted output of the computation is being sent to the legitimate data user.

3. The data user decrypts and gets the output of the computation. It does not learn the original values

E.g. , Linear regressive models are used in patient monitoring, in predicting medical device performance etc. Such computations can be performed over homomorphically encrypted values

1. The data holders send encrypted data to the intermediary service; only the data user can decrypt.

Homomorphic encryption



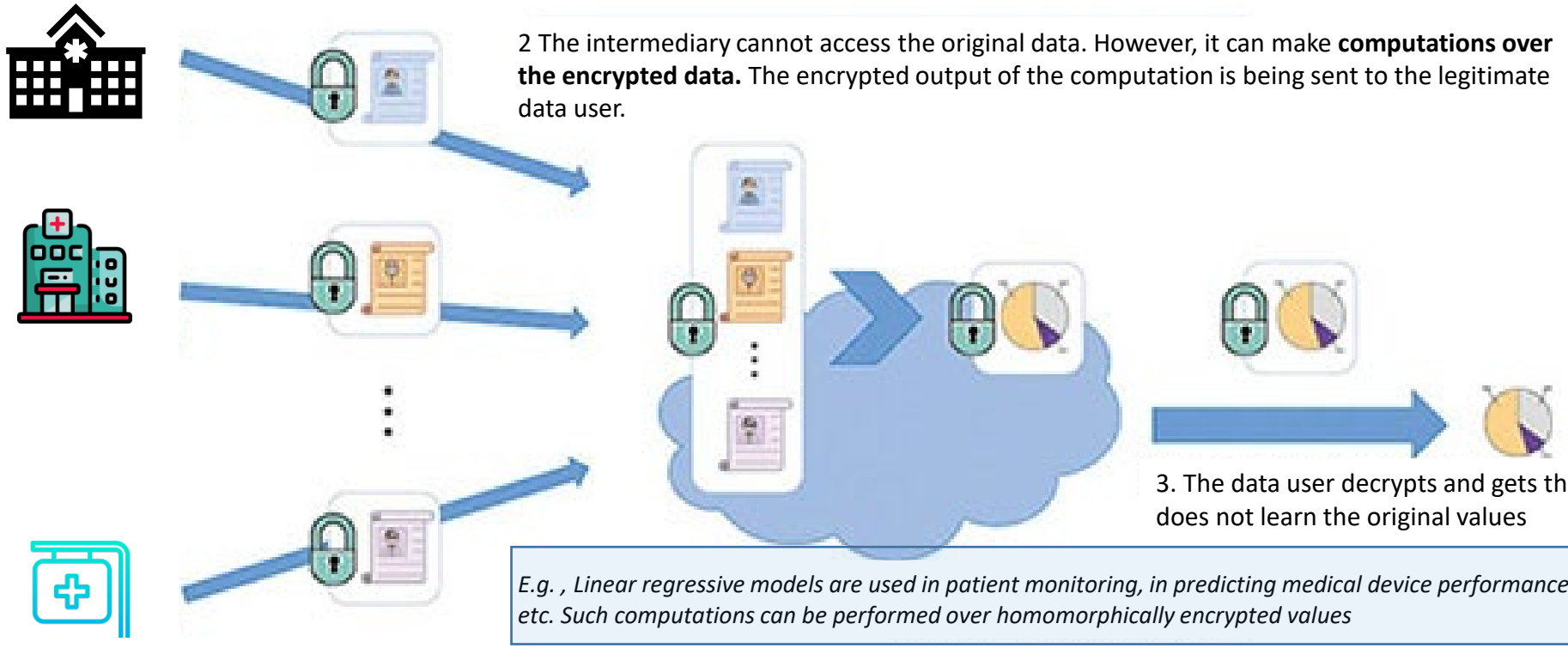
2 The intermediary cannot access the original data. However, it can make **computations over the encrypted data.** The encrypted output of the computation is being sent to the legitimate data user.

3. The data user decrypts and gets the output of the computation. It does not learn the original values

E.g. , Linear regressive models are used in patient monitoring, in predicting medical device performance etc. Such computations can be performed over homomorphically encrypted values

1. The data holders send encrypted data to the intermediary service; only the data user can decrypt.

Homomorphic encryption



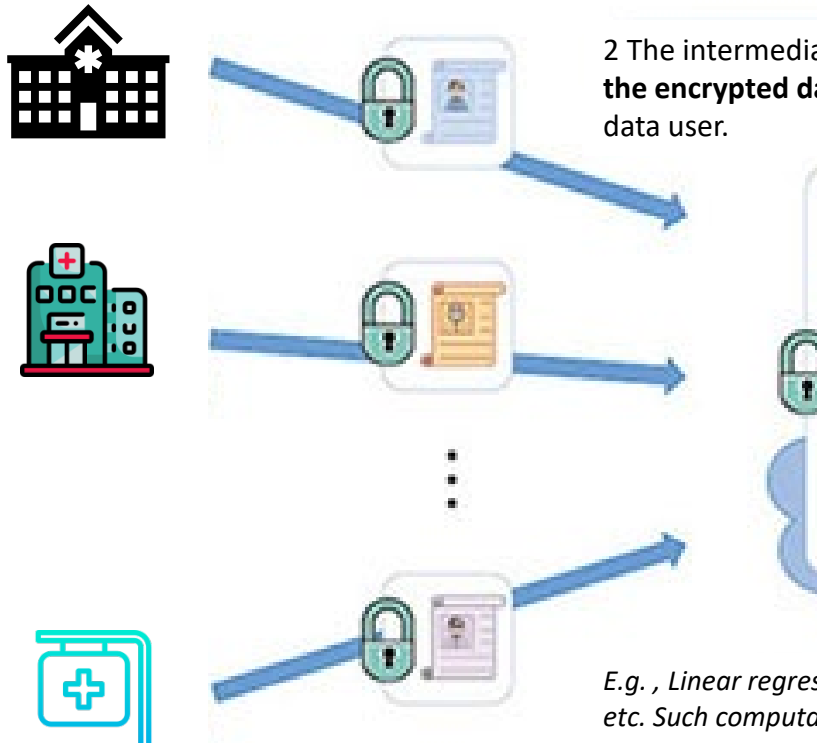
2 The intermediary cannot access the original data. However, it can make **computations over the encrypted data.** The encrypted output of the computation is being sent to the legitimate data user.

3. The data user decrypts and gets the output of the computation. It does not learn the original values

E.g. , Linear regressive models are used in patient monitoring, in predicting medical device performance etc. Such computations can be performed over homomorphically encrypted values

1. The data holders send encrypted data to the intermediary service; only the data user can decrypt.

Homomorphic encryption



2 The intermediary cannot access the original data. However, it can make **computations over the encrypted data.** The encrypted output of the computation is being sent to the legitimate data user.

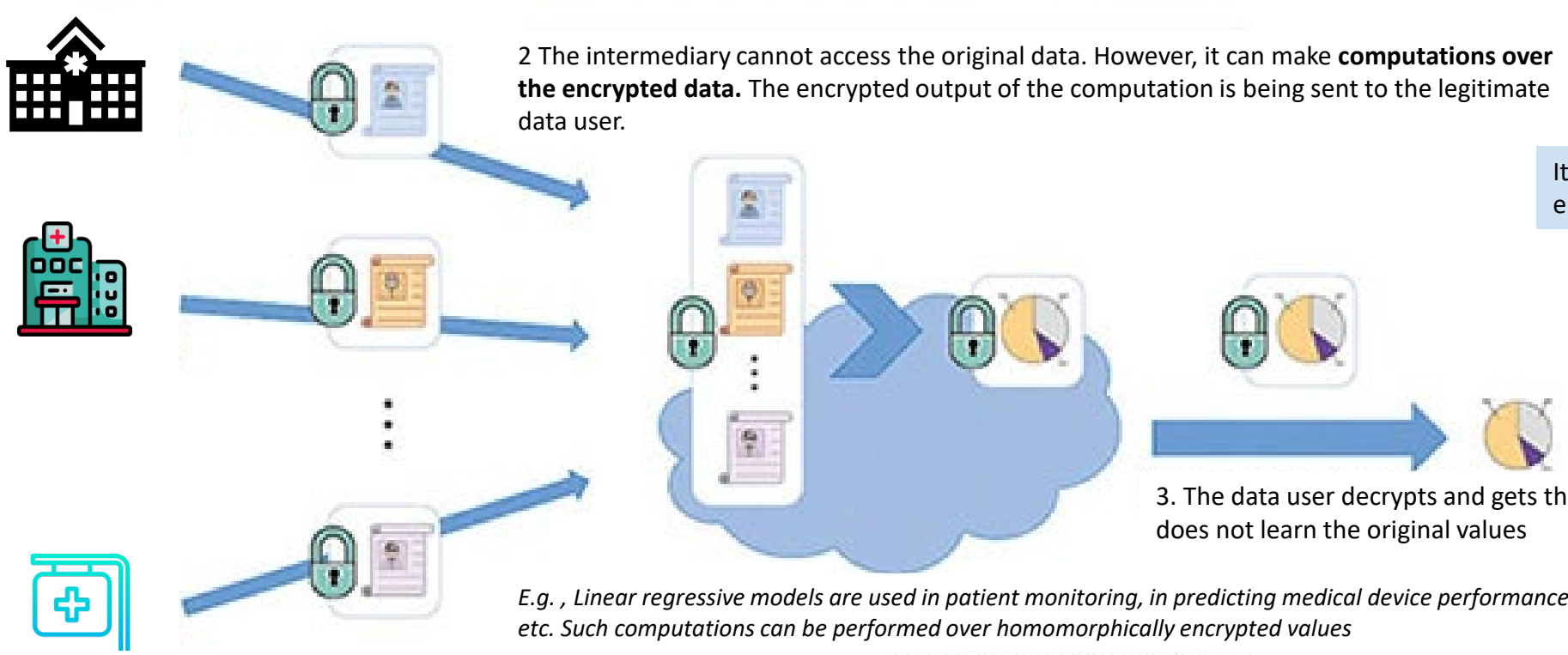
3. The data user decrypts and gets the output of the computation. It does not learn the original values

E.g. , Linear regressive models are used in patient monitoring, in predicting medical device performance etc. Such computations can be performed over homomorphically encrypted values

1. The data holders send encrypted data to the intermediary service; only the data user can decrypt.

Homomorphic encryption

Available libraries for homomorphic encryption – e.g. Zama – (https://www.zama.ai/)



2 The intermediary cannot access the original data. However, it can make **computations over the encrypted data.** The encrypted output of the computation is being sent to the legitimate data user.

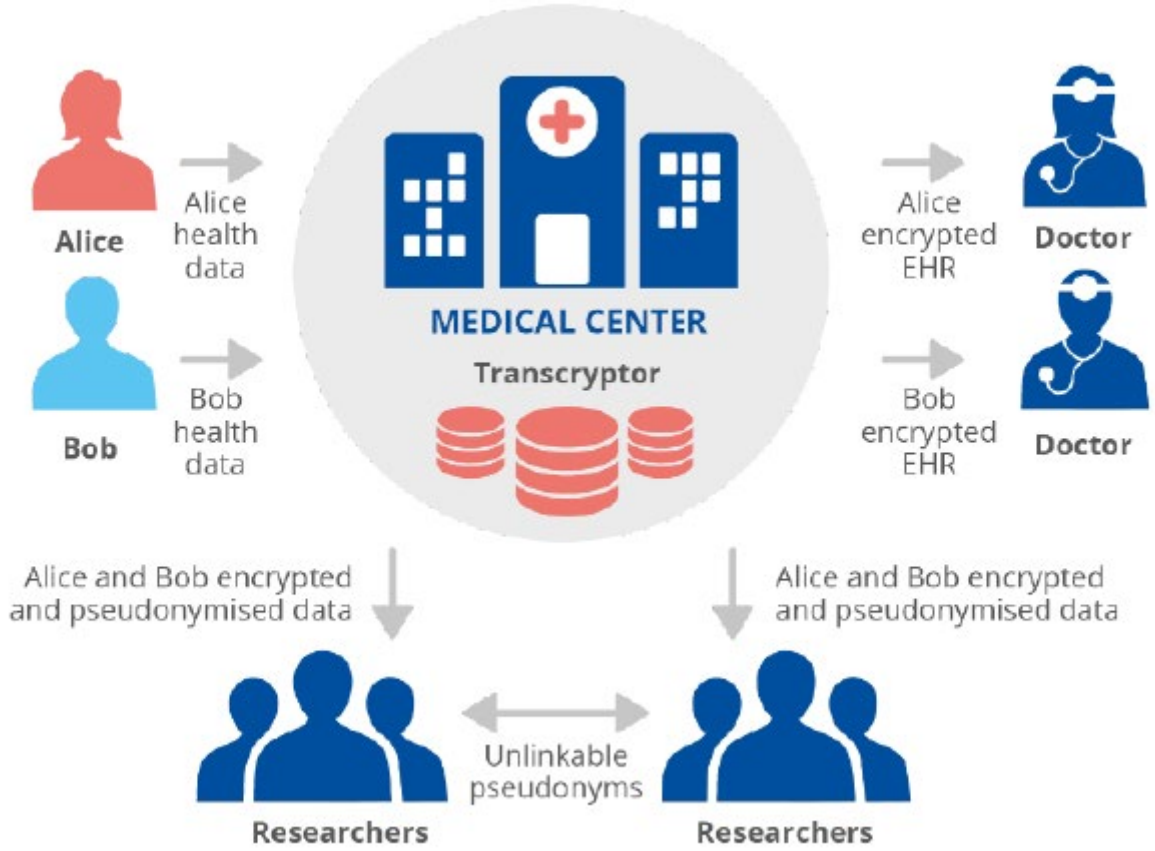
It can be combined with proxy re-encryption

3. The data user decrypts and gets the output of the computation. It does not learn the original values

E.g. , Linear regressive models are used in patient monitoring, in predicting medical device performance etc. Such computations can be performed over homomorphically encrypted values

1. The data holders send encrypted data to the intermediary service; only the data user can decrypt.

Polymorphic encryption

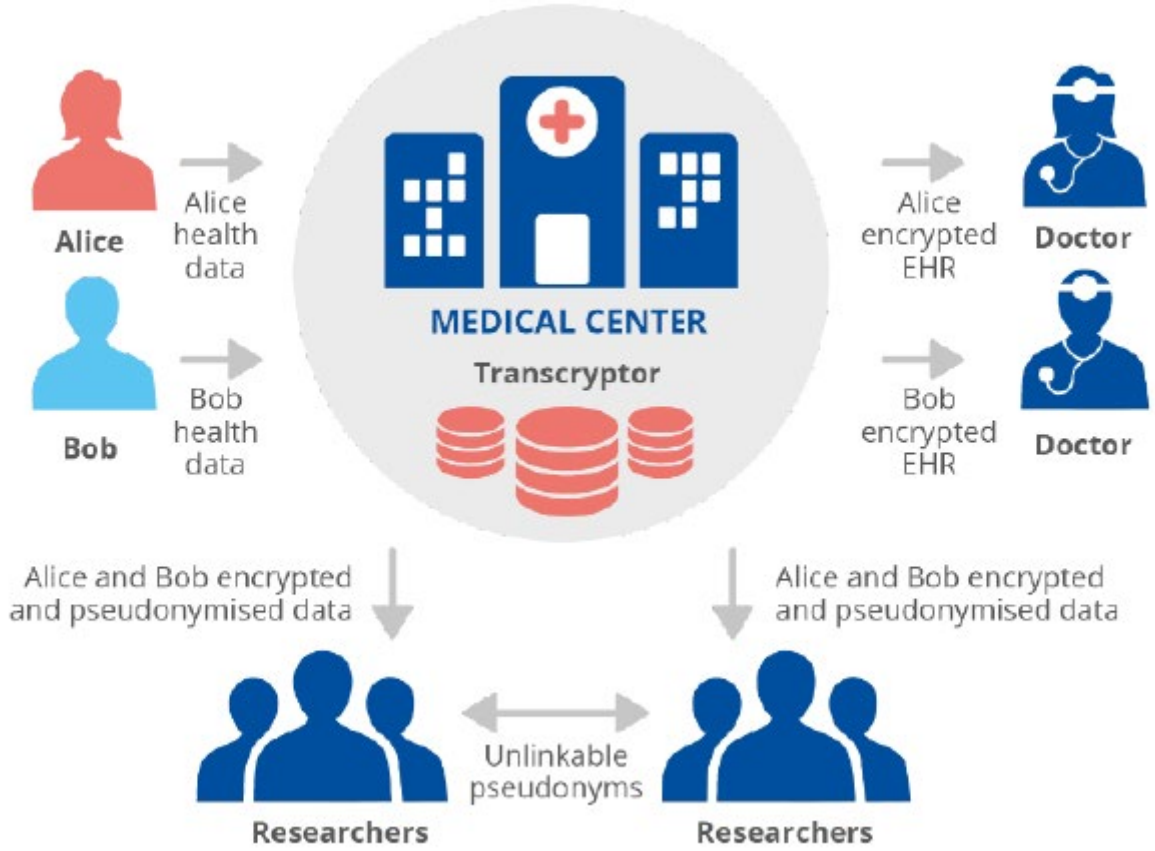


Source: ENISA, Data Sharing Report, 2023

- The data are being stored encrypted, without having knowledge of who will request to decrypt them
- If a data user requests access, then the data will be appropriately re-encrypted so as only this user will be able to decrypt
- Simultaneously, the data can be also pseudonymised, deriving unlinkable pseudonyms between different data users.
- The entity performing these tasks (transcryptor) does not have access to the original plaintext data (i.e. these tasks are being performed "blindly")
- The transcryptor could be the main entity implementing a SPE!
- This approach has been already used

B. Gastel et. al., **Data Protection Using Polymorphic Pseudonymisation in a Large-Scale Parkinson's Disease Study,** 2021

Polymorphic encryption

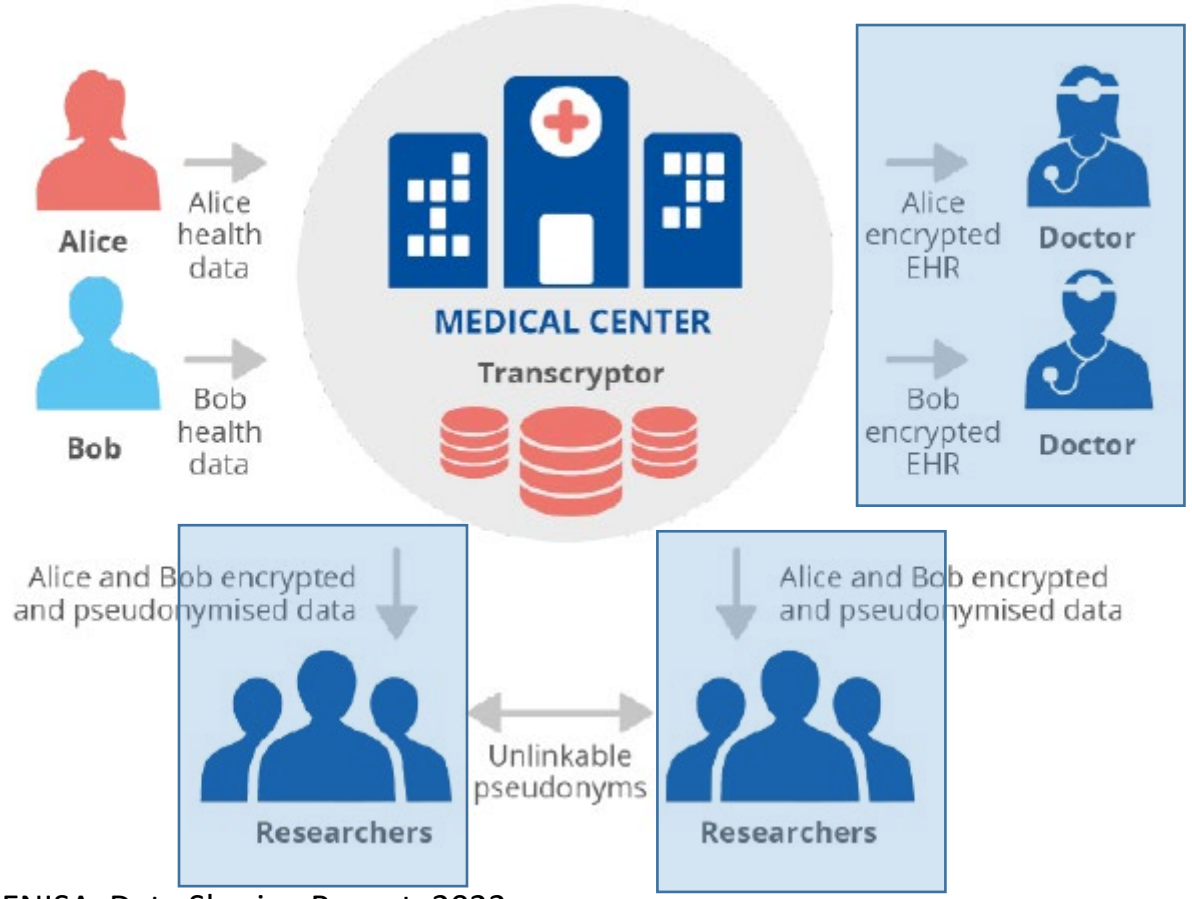


Source: ENISA, Data Sharing Report, 2023

- The data are being stored encrypted, without having knowledge of who will request to decrypt them
- If a data user requests access, then the data will be appropriately re-encrypted so as only this user will be able to decrypt
- Simultaneously, the data can be also pseudonymised, deriving unlinkable pseudonyms between different data users.
- The entity performing these tasks (transcryptor) does not have access to the original plaintext data (i.e. these tasks are being performed "blindly")
- The transcryptor could be the main entity implementing a SPE!
- This approach has been already used

B. Gastel et. al., **Data Protection Using Polymorphic Pseudonymisation in a Large-Scale Parkinson's Disease Study,** 2021

Polymorphic encryption

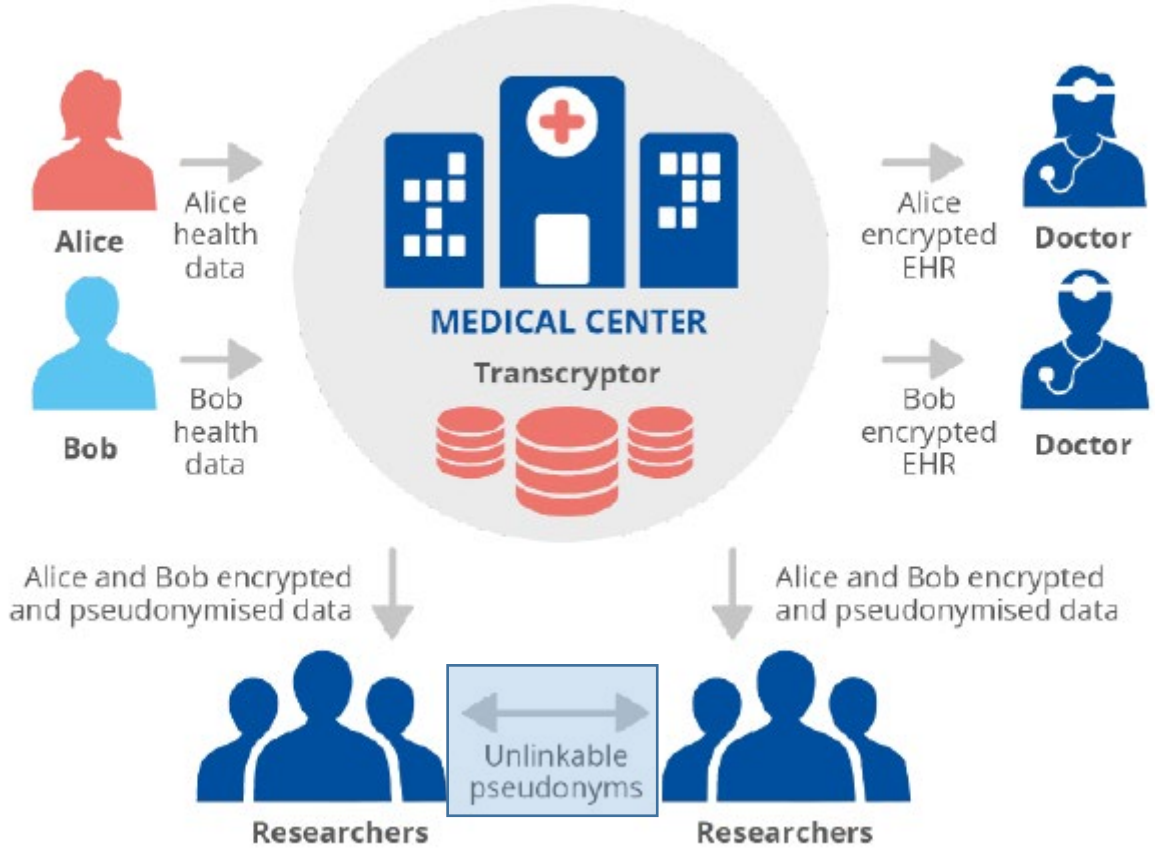


Source: ENISA, Data Sharing Report, 2023

- The data are being stored encrypted, without having knowledge of who will request to decrypt them
- If a data user requests access, then the data will be appropriately re-encrypted so as only this user will be able to decrypt
- Simultaneously, the data can be also pseudonymised, deriving unlinkable pseudonyms between different data users.
- The entity performing these tasks (transcryptor) does not have access to the original plaintext data (i.e. these tasks are being performed "blindly")
- The transcryptor could be the main entity implementing a SPE!
- This approach has been already used

B. Gastel et. al., **Data Protection Using Polymorphic Pseudonymisation in a Large-Scale Parkinson's Disease Study,** 2021

Polymorphic encryption

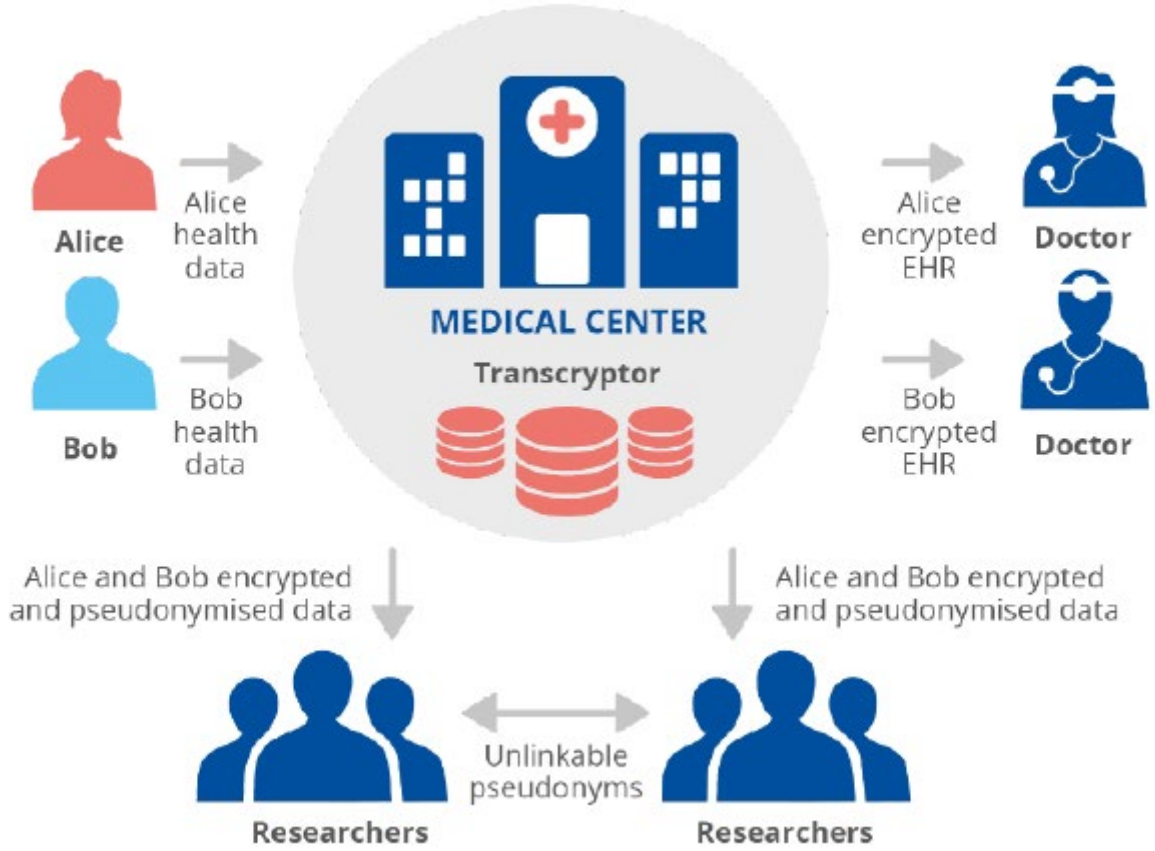


Source: ENISA, Data Sharing Report, 2023

- The data are being stored encrypted, without having knowledge of who will request to decrypt them
- If a data user requests access, then the data will be appropriately re-encrypted so as only this user will be able to decrypt
- Simultaneously, the data can be also pseudonymised, deriving unlinkable pseudonyms between different data users.
- The entity performing these tasks (transcryptor) does not have access to the original plaintext data (i.e. these tasks are being performed "blindly")
- The transcryptor could be the main entity implementing a SPE!
- This approach has been already used

B. Gastel et. al., **Data Protection Using Polymorphic Pseudonymisation in a Large-Scale Parkinson's Disease Study,** 2021

Polymorphic encryption

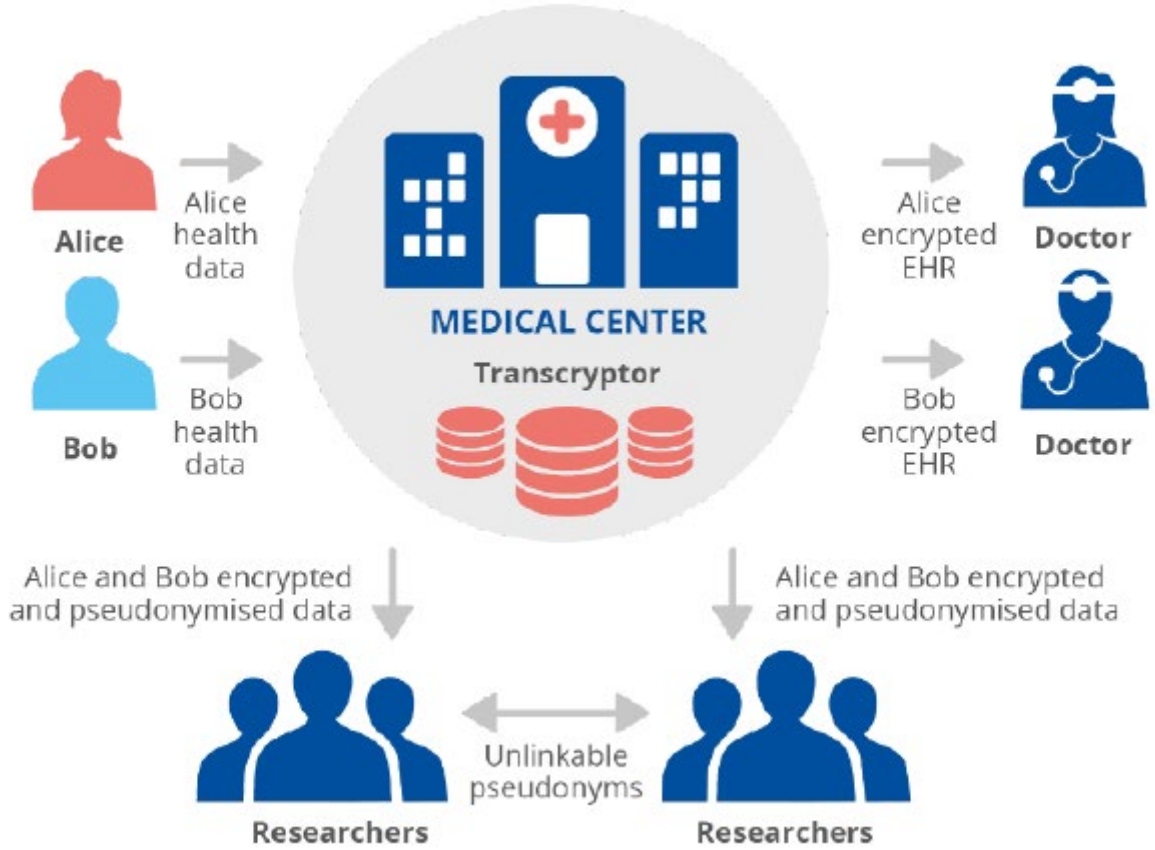


Source: ENISA, Data Sharing Report, 2023

- The data are being stored encrypted, without having knowledge of who will request to decrypt them
- If a data user requests access, then the data will be appropriately re-encrypted so as only this user will be able to decrypt
- Simultaneously, the data can be also pseudonymised, deriving unlinkable pseudonyms between different data users.
- The entity performing these tasks (transcryptor) does not have access to the original plaintext data (i.e. these tasks are being performed "blindly")
- The transcryptor could be the main entity implementing a SPE!
- This approach has been already used

B. Gastel et. al., **Data Protection Using Polymorphic Pseudonymisation in a Large-Scale Parkinson's Disease Study,** 2021

Polymorphic encryption

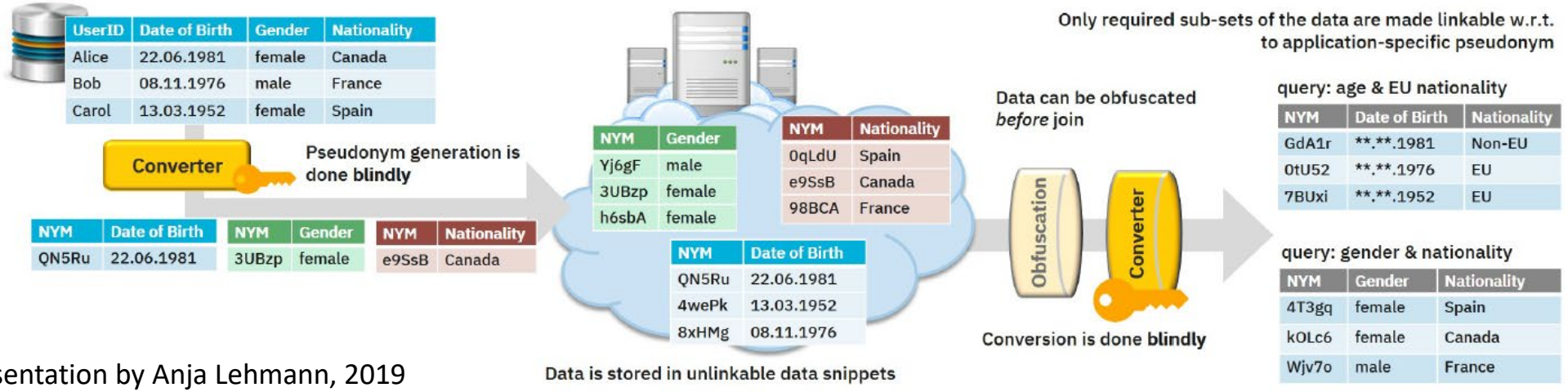


Source: ENISA, Data Sharing Report, 2023

- The data are being stored encrypted, without having knowledge of who will request to decrypt them
- If a data user requests access, then the data will be appropriately re-encrypted so as only this user will be able to decrypt
- Simultaneously, the data can be also pseudonymised, deriving unlinkable pseudonyms between different data users.
- The entity performing these tasks (transcryptor) does not have access to the original plaintext data (i.e. these tasks are being performed "blindly")
- The transcryptor could be the main entity implementing a SPE!
- This approach has been already used

  B. Gastel et. al., **Data Protection Using Polymorphic Pseudonymisation in a Large-Scale Parkinson's Disease Study,** 2021

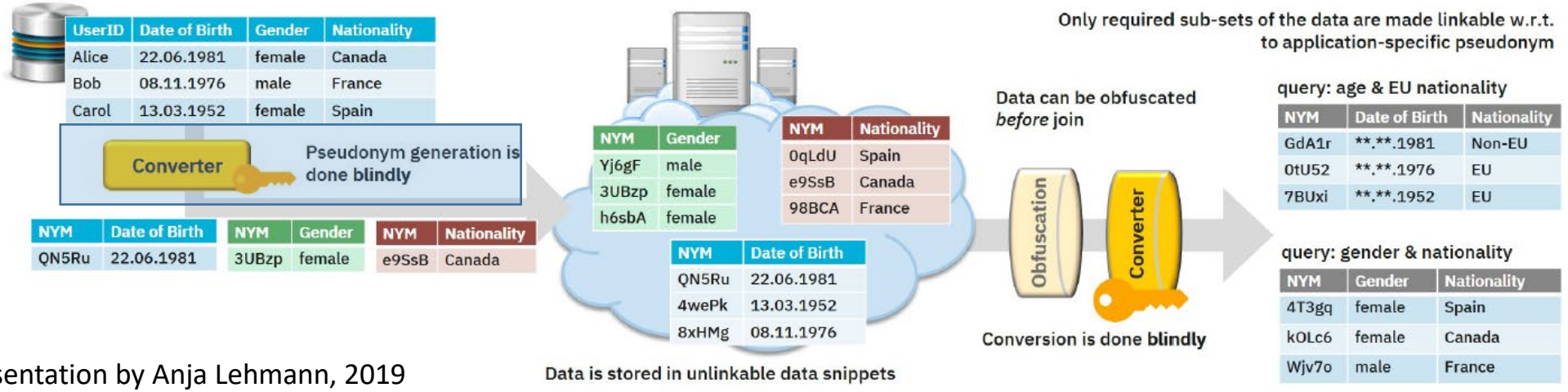
Oblivious pseudonymisation



Source: Presentation by Anja Lehmann, 2019

- The entity generating the first pseudonyms (converter) does not get access to the original identifiers
- The intermediary stores and processes pseudonymous data in several unlinkable tables, without being able to re-identify the individuals . It does not get access to the original identifies.
- Upon a legitimate data user's request, the data can be combined from the different tables and re-pseudonymised.
  - Different – and unlinkable - pseudonyms for different data users
  - The intermediary **does never learn the original identifiers** .

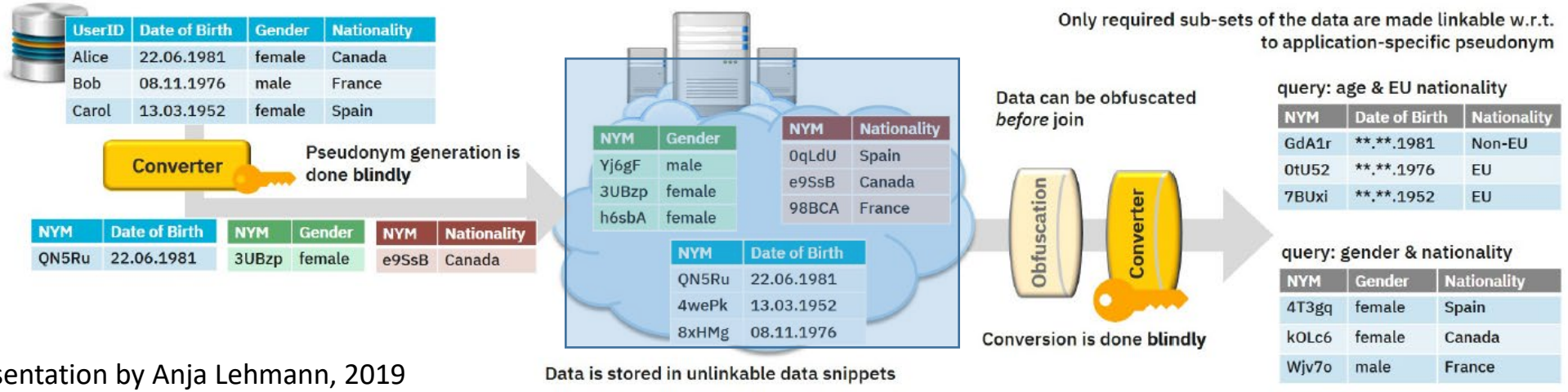
Oblivious pseudonymisation



Source: Presentation by Anja Lehmann, 2019

- The entity generating the first pseudonyms (converter) does not get access to the original identifiers
- The intermediary stores and processes pseudonymous data in several unlinkable tables, without being able to re-identify the individuals . It does not get access to the original identifies.
- Upon a legitimate data user's request, the data can be combined from the different tables and re-pseudonymised.
  - Different – and unlinkable - pseudonyms for different data users
  - The intermediary **does never learn the original identifiers** .

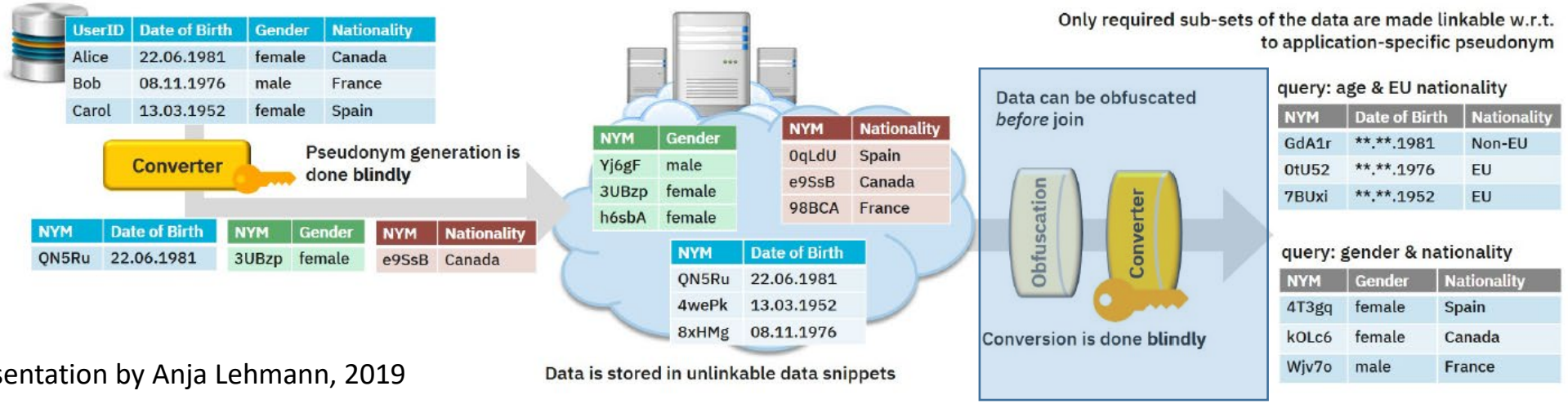
Oblivious pseudonymisation



Source: Presentation by Anja Lehmann, 2019

- The entity generating the first pseudonyms (converter) does not get access to the original identifiers
- The intermediary stores and processes pseudonymous data in several unlinkable tables, without being able to re-identify the individuals . It does not get access to the original identifies.
- Upon a legitimate data user's request, the data can be combined from the different tables and re-pseudonymised.
  - Different – and unlinkable - pseudonyms for different data users
  - The intermediary **does never learn the original identifiers** .

Oblivious pseudonymisation



Source: Presentation by Anja Lehmann, 2019

- The entity generating the first pseudonyms (converter) does not get access to the original identifiers
- The intermediary stores and processes pseudonymous data in several unlinkable tables, without being able to re-identify the individuals . It does not get access to the original identifies.
- Upon a legitimate data user's request, the data can be combined from the different tables and re-pseudonymised.
  - Different – and unlinkable - pseudonyms for different data users
  - The intermediary **does never learn the original identifiers** .

Summary - Conclusions

- (Any type of) intermediaries may collect huge amount of (sensitive) personal data
  - Including intermediaries as defined in the DGA, as well as the Health Data Access Body in the EHDS
- A SPE is necessary but, ….. how to ensure that all the desired properties are in place?
- Advanced cryptographic techniques seem to address some data protection requirements
  - Not a panacea, but they should be taken into account, in the context of a risk-based approach
  - Their implementations are feasible…
- Need to ensure that such techniques will be further investigated/promoted
  - How? (through developing standards?)

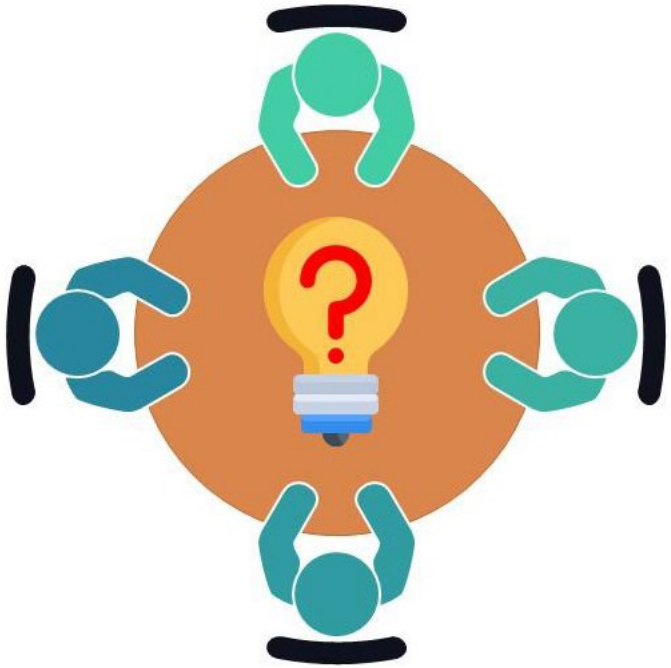
Useful sources

- ENISA, “Engineering personal data protection in EU Data Spaces”, 2024.
- ENISA, “Engineering personal data sharing – Emerging use cases and technologies”, 2023.
- M.R. Albrecht, A. Davidson, A. Deo, D. Gardham, “Crypto Dark Matter on the Torus: Oblivious PRFs from shallow PRFs and TFHE”. EUROCRYPT 2024.
- A. Lehmann, “ScrambleDB: Oblivious (Chameleon) Pseudonymization-as-a-Service“, Proc. of PETs Symposium, 2019
- Eric Verheul, Bart Jacobs, Carlo Meijer, Mireille Hildebrandt, and Joeri de Ruiter, “Polymorphic encryption and pseudonymisation”, Cryptology ePrint Archive, Report 2016/411, 2016.
- B. Gastel, B. Jacobs, J. Popma, “Data Protection Using Polymorphic Pseudonymisation in a Large-Scale Parkinson’s Disease Study”, 2021
- K. Limniotis, “Cryptography as the means to protect fundamental human rights‘’, Cryptography, vol. 5, no. 4, 2021.

«Cryptography is about the right to privacy, freedom of speech, freedom of political association, freedom of the press, freedom from unreasonable search and seizure, freedom to be left alone».

Phil Zimmermann

Thank you for your attention!



Discussion / Questions ?