# Data Sharing  - Use cases in the Health Care Sector

Dr. Konstantinos Limniotis

# Presentation overview

- Challenges in data sharing in the health sector
- Use case scenarios
    1. User-controlled data sharing
    2. Sharing electronic health records for medical and/or research purposes
    3. Finding common patients between health centers
- Usage of (advanced) data protection engineering
    - Why “conventional” technical solutions may not be adequate
- Concluding remarks

# The need for sharing health data

- Why is it important?
  - Strengthen coordination and collaboration between the public and private health care entities towards providing
    - Effective personalised health-care assistance
    - Achieving public health goals
    - Conducting scientific research (including clinical trials)

Picture from freepik.com

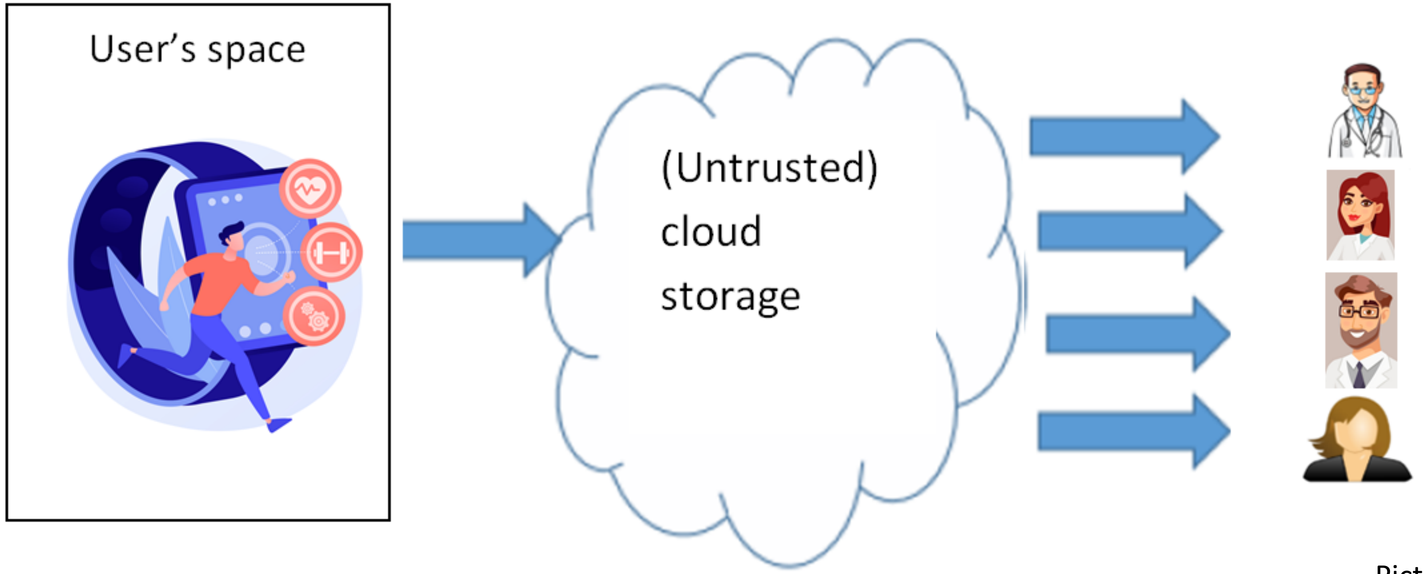
# Sharing health data – Main challenges

- High data protection risks to the sensitivity and volume of such data
- Not always easy to ensure the fulfilment of data protection principles such as transparency and data minimisation
  - Multiple sources of patient data
  - Linking different datasets may be technically easy
- Requirements due to specific (national) legal obligations

➢ A cautious implementation of the "data protection by design" principle is essential

  ➢ Data protection engineering may be the vehicle to support this principle

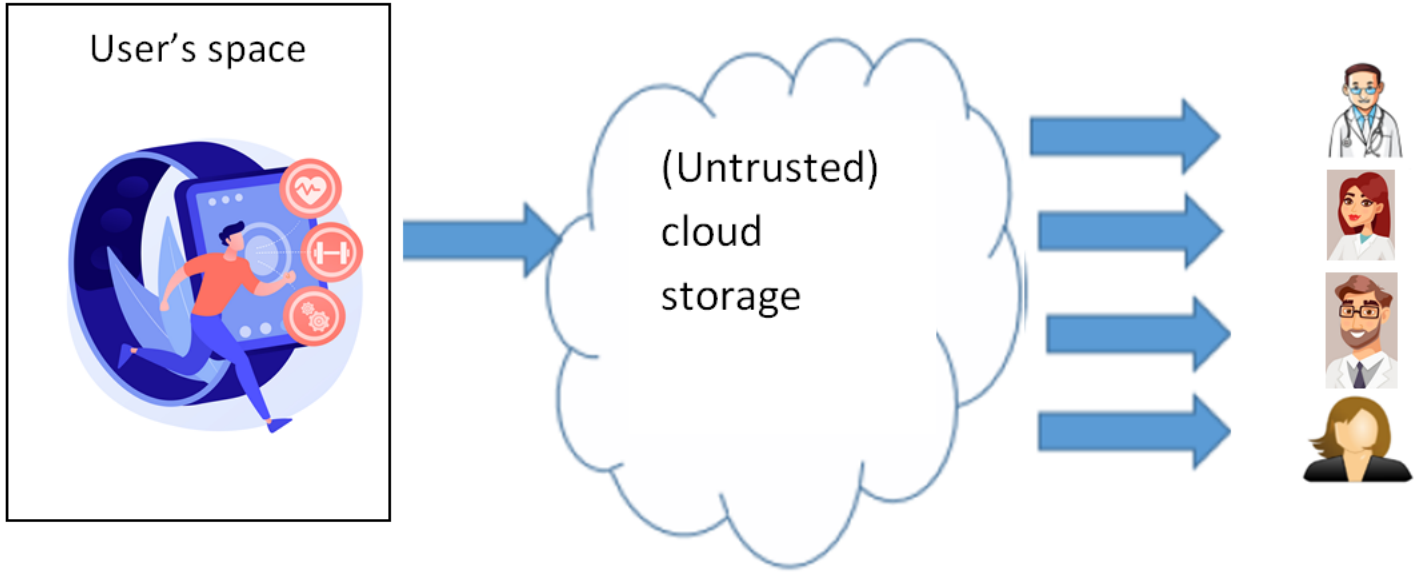
# Use case scenario 1 - User-controlled data sharing




Pictures from freepik.com

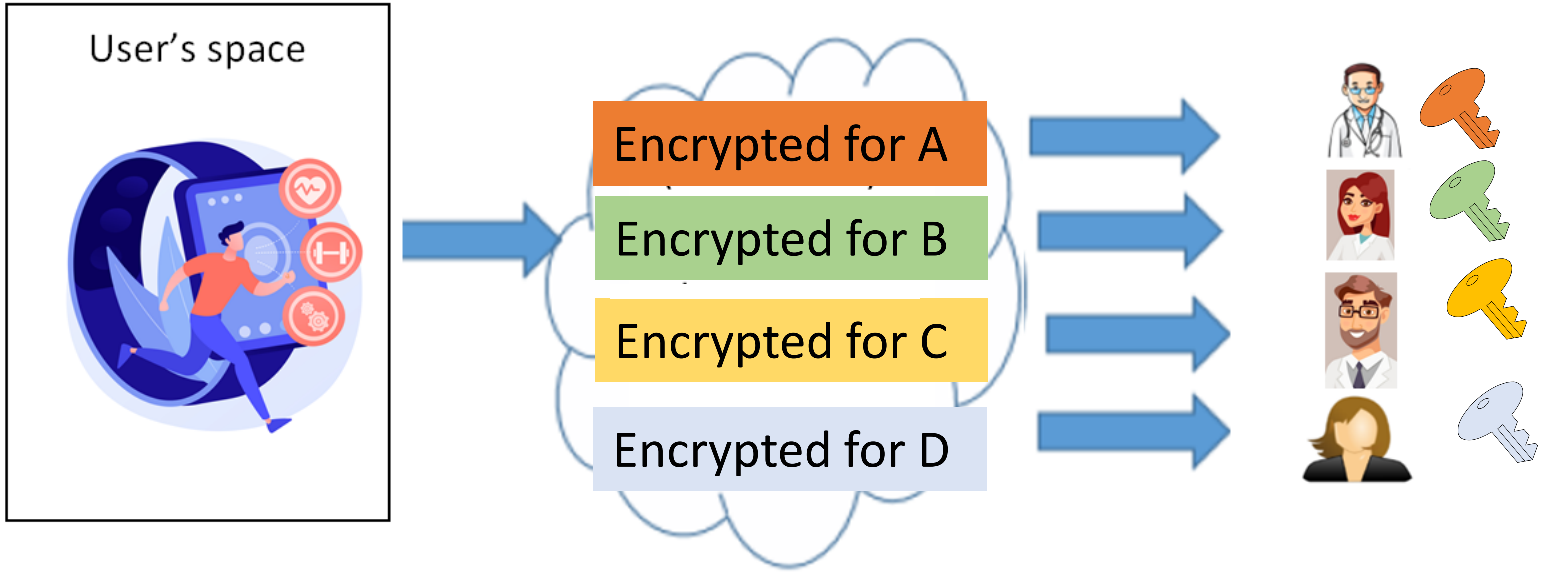
- The user uses a wearable device for, e.g., continuous glucose monitoring (CGM),which also monitors blood pressure, caffeine levels and lactate levels.
- The user uploads them to the cloud with the aim to be subsequently accessed, at any time, by the user herself, as well as by other entities – e.g. by her personal doctors.

# Basic requirements for ensuring user-controlled data sharing




- The cloud provider is "untrusted"; it should not be able to read/manipulate the original data
- The user should be able to securely and selectively share the data streams generated by her device, in a dynamic access model
  - E.g. Doctor X may get access only to those data that correspond to the last three months and/or only only for data corresponding to user's blood pressure

# Limitations of "classical" cryptography

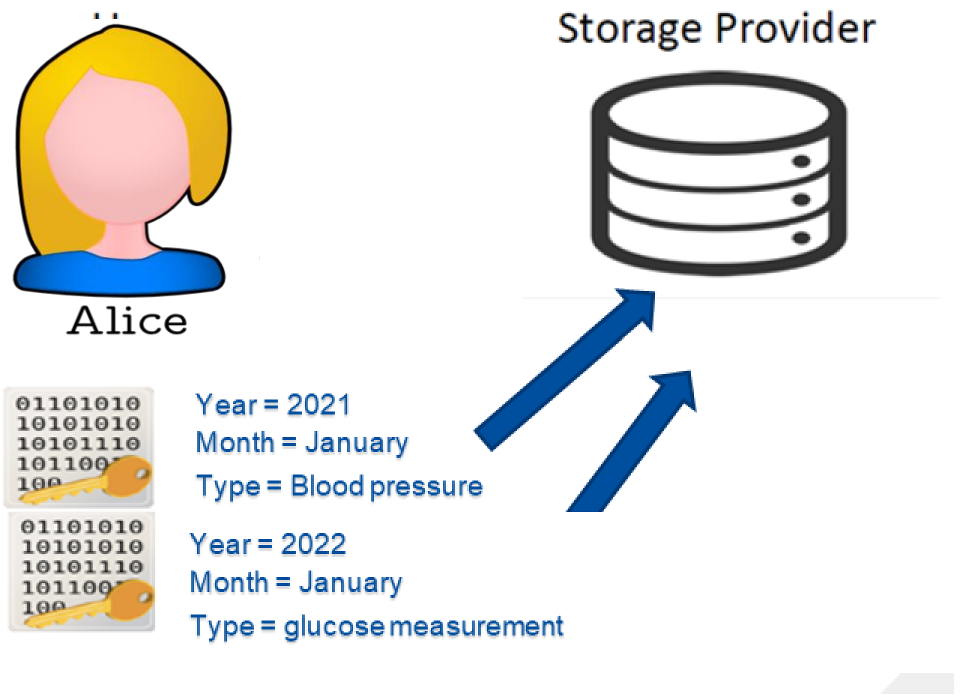



- If the same data are to be shared with multiple entities, the user needs to store the same data many times, each encrypted for each entity
- The access permission is actually determined a priori in the encryption time
  - But the desired recipients may not be known in advance
  - ➢ For each new required access, a new encryption is needed

# A way to proceed: Attribute-based encryption (ABE)

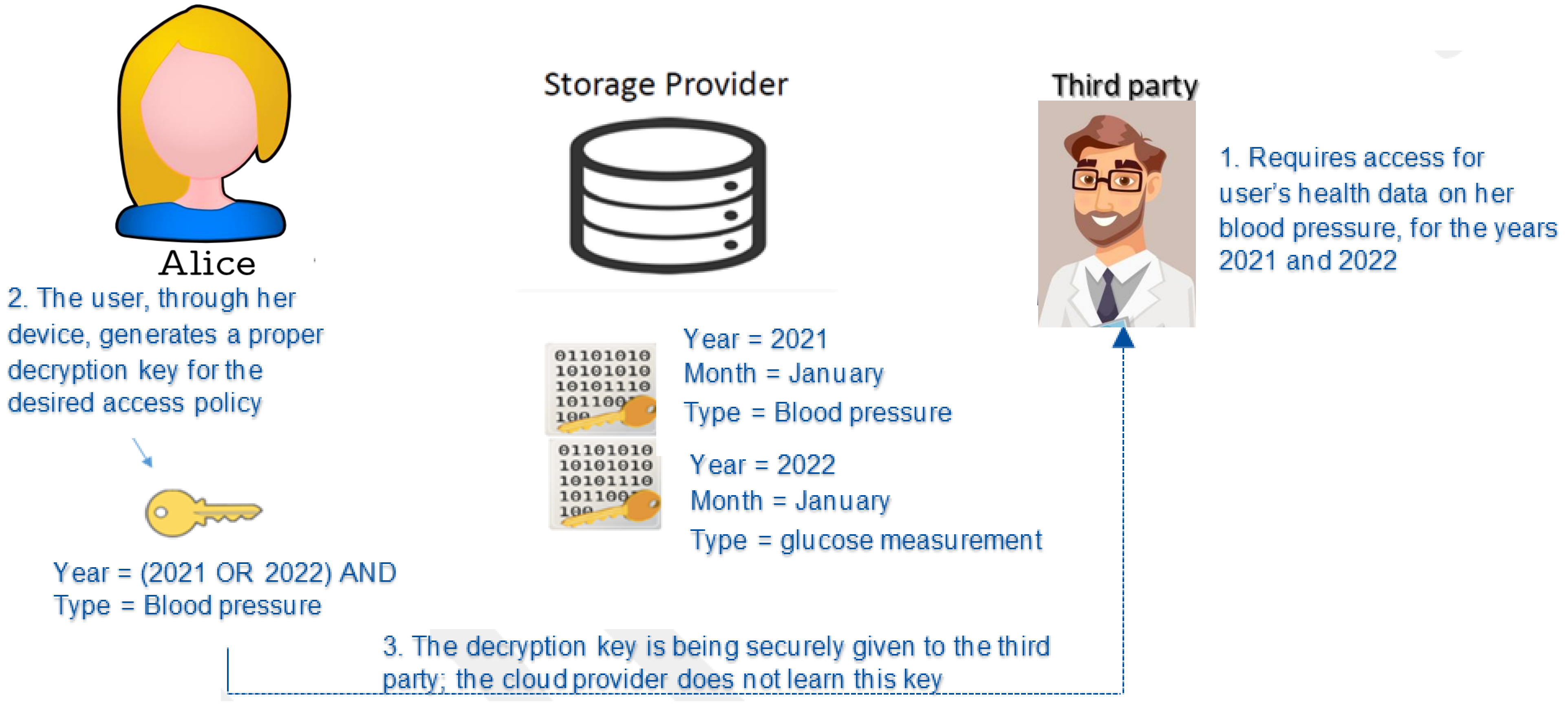
- ABE is a special case of public-key encryption
- Data can be encrypted by a public key but, at the same time, there may be more than one decryption keys
  - Each of them depending on other small pieces of additional information related with the data, being called attributes (e.g. “type=blood pressure” ).
- Access policies can be defined according to the attributes defined such as, e.g. “type=blood pressure” AND “year>=2021”.
  - Decryption keys are being generated according to such access policies
  - The owner of the decryption key can decrypt only this part of the data that satisfies the corresponding access policy

# Revisiting our scenario: ABE in practice

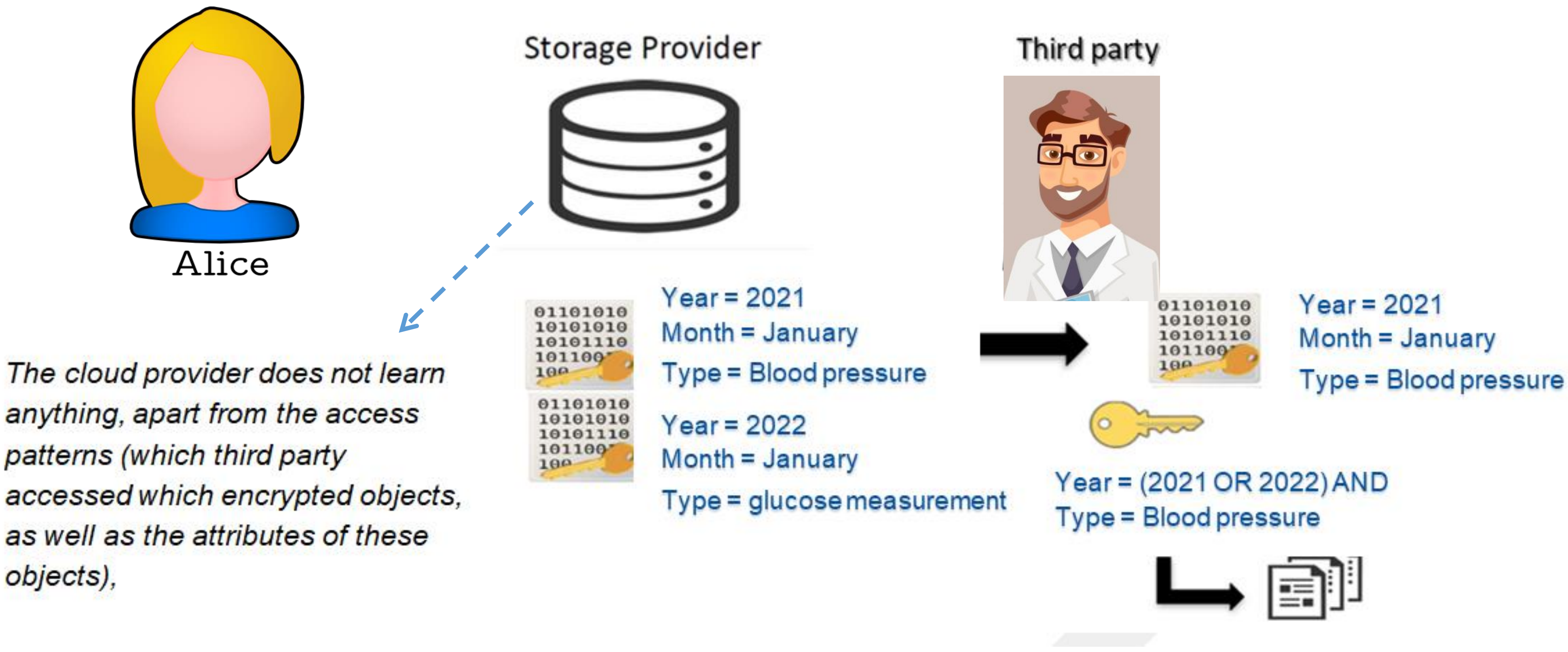



- The user (through her device) assigns tags to the relevant objects based on specific attributes.
- The data are being encrypted through ABE encryption and uploaded to the cloud provider
  - Decryption key is not available to the cloud

# Revisiting our scenario: ABE in practice (*Cont.*)



# Revisiting our scenario: ABE in practice (*Cont.)*



# Discussion so far…

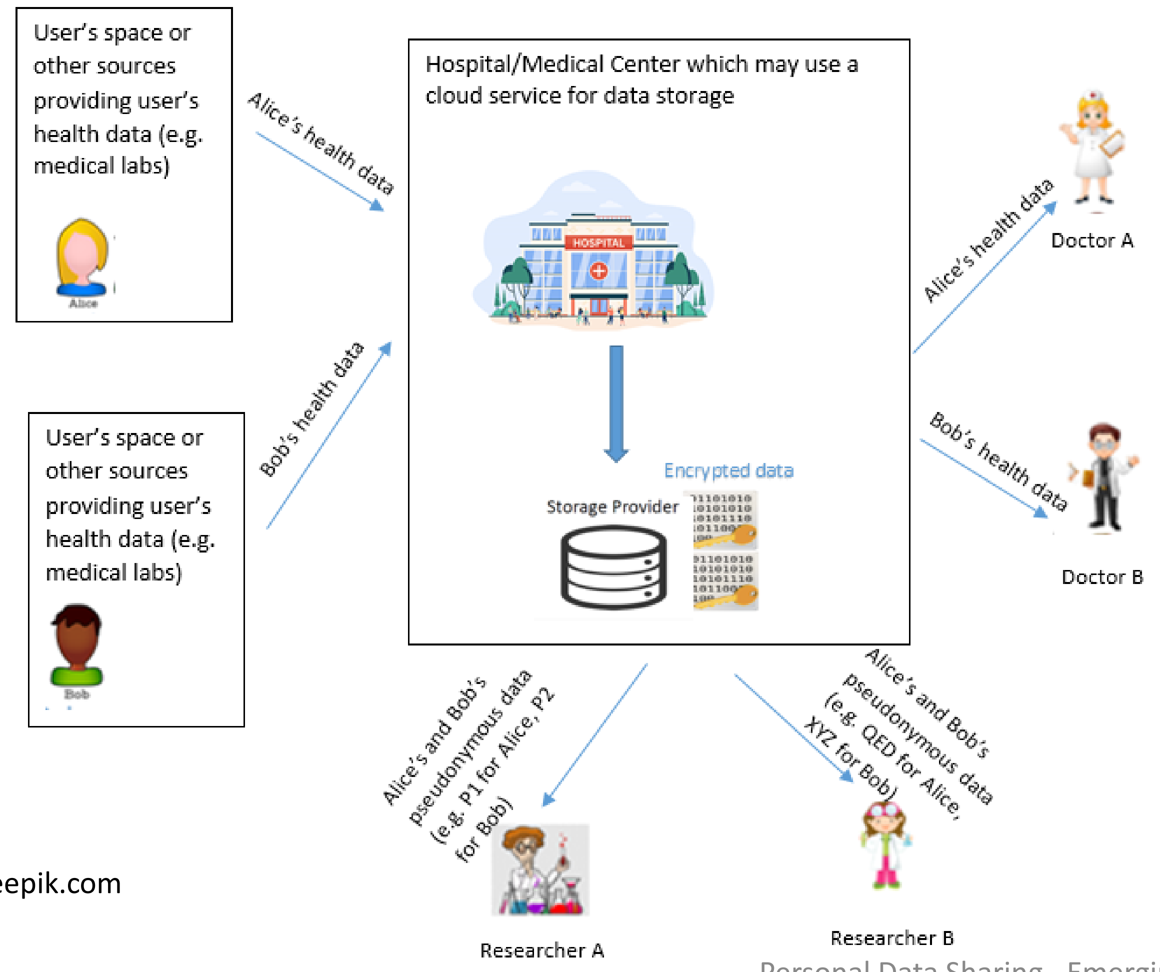
- Advanced cryptographic techniques allow for user-controlled data sharing between two entities such as:
  - The storage facility is not able to decrypt
  - Dynamically determine (after the encryption) who can get access to the data and for which data

➢ Other approaches in this direction: Proxy re-encryption

➢ They may yield proper implementations for ensuring user's consent when this is the proper legal basis for the processing

➢ They can also be applied in other data sharing scenarios

# Use case scenario 2 - Sharing health records for medical and/or research purposes




Pictures from freepik.com

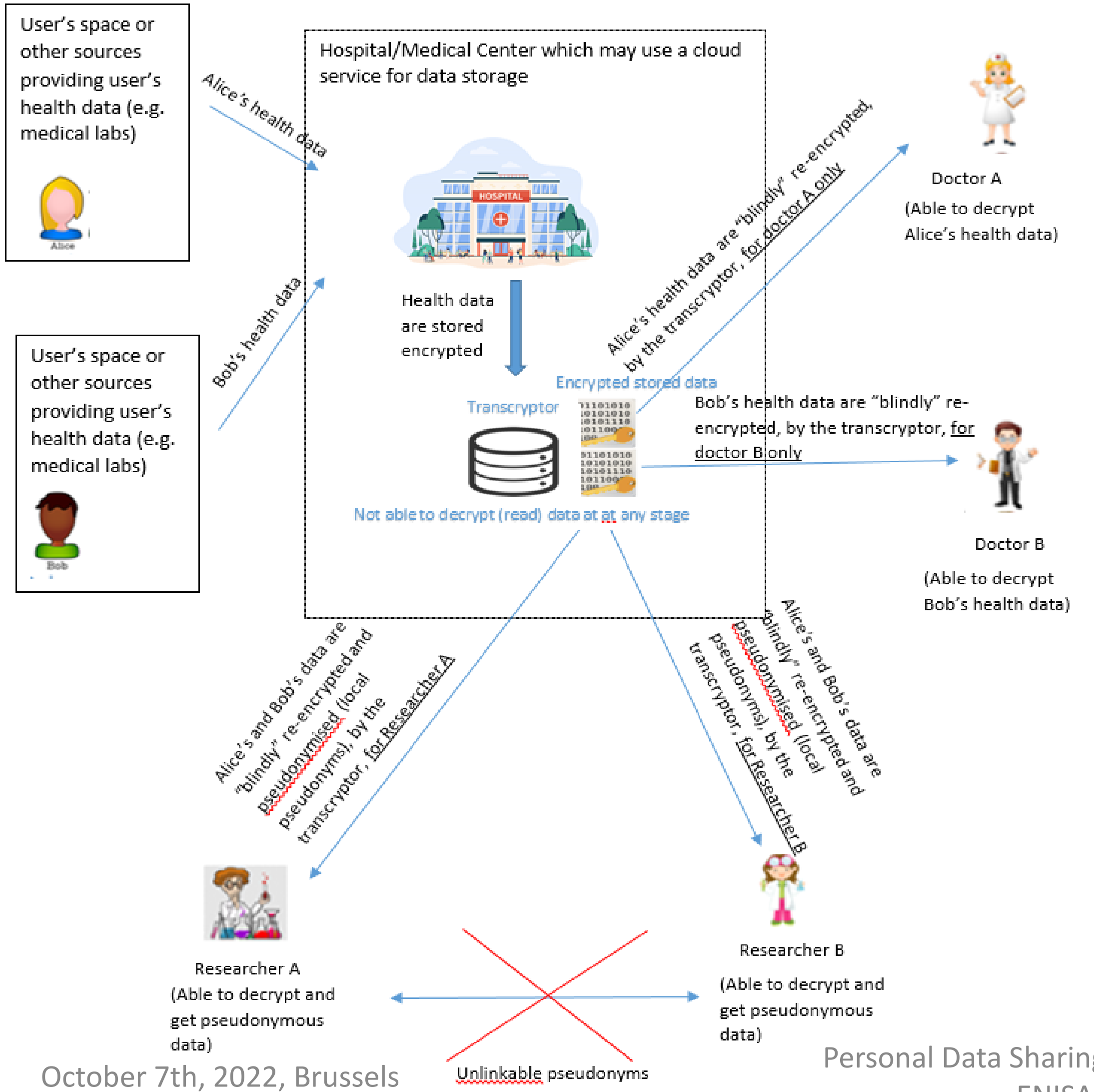
**Basic requirements**

- The storage provider may be internal or a cloud service, being "untrusted"
- Doctors should securely receive personalised health data, for specific patients, to provide health services
- Researchers should receive pseudonymous data (research-oriented), being irreversible for them and unlinkable with other pseudonymous data stemming from the same data set

# A way to proceed: Polymorphic Encryption and Pseudonymisation (PEP)

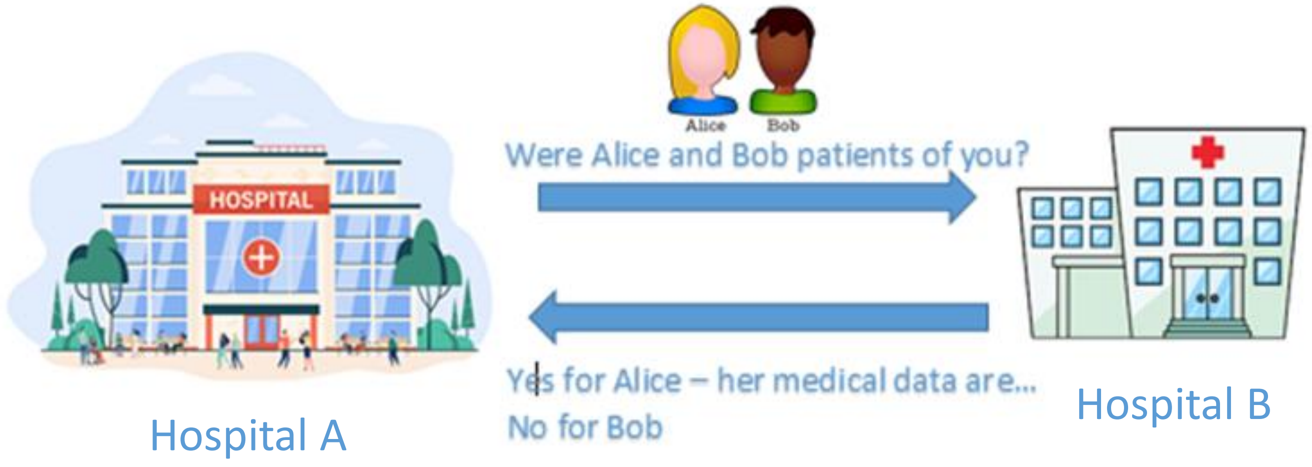
- Encryption: Data can be encrypted (and stored at a central point – e.g. at a cloud service) in such a way that there is no need to fix a priori who can decrypt the data later;
  - This can be decided later on, via some transformation of the ciphertext
  - This transformation can be performed blindly, without the party performing this (the transcryptor) being able to read the original data
- Pseudonymisation: Pseudonyms are being cryptographically generated over the encrypted data
  - The transcryptor can also "change" the content of the ciphertext so as the original user's identifier in the original data is transformed into a suitable pseudonym
  - Again, this takes place blindly (i.e. over the unintelligible encrypted data)

# PEP in practice




- The transcryptor is able to "transform" the ciphertext for any possible recipient, so as only this recipient can decrypt
  - And, depending on the case, the data may be decrypted in a pseudonymised form
- The transcryptor operates blindly (not having access to the original data)
- A relatively new approach, with one active application.
- The corresponding research team also envisions PEP in a user-controlled data sharing model.

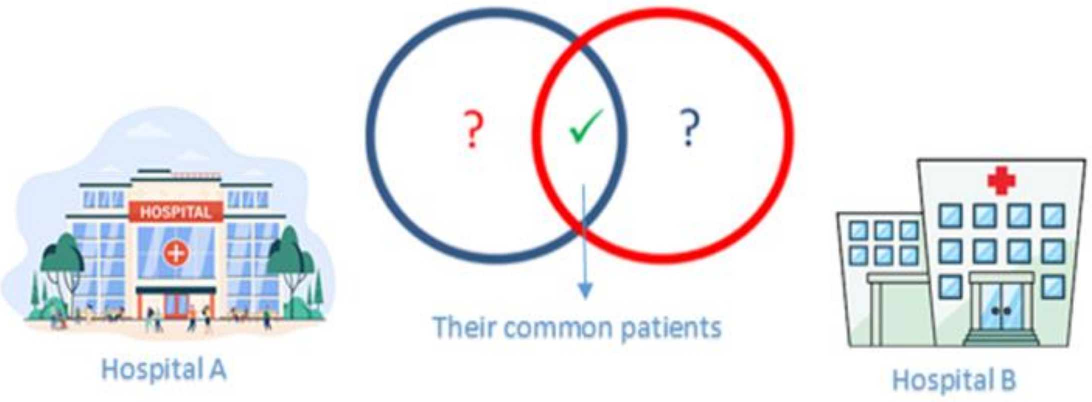
# Use case scenario 3 - Finding common patients between two health centers




Pictures from freepik.com

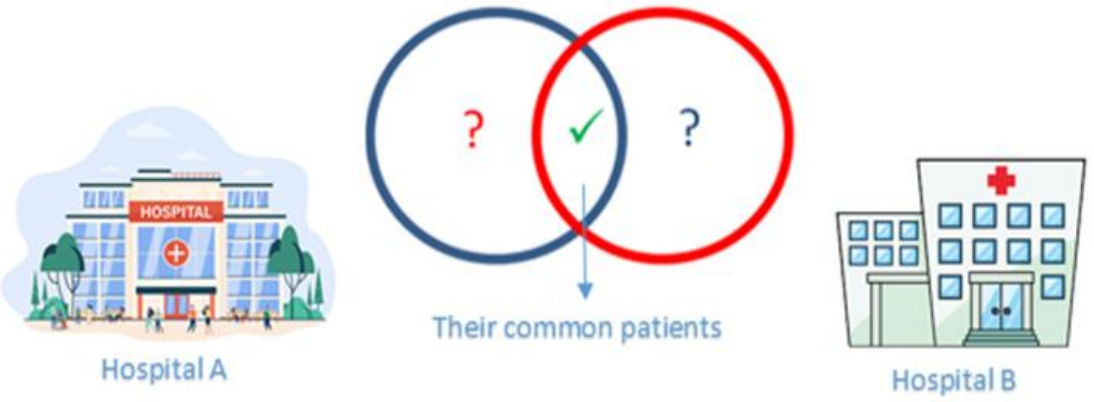
- The hospital A wants to learn if some of its patients have been previously hospitalized in the hospital B.
- If yes, their corresponding health data should be provided by B.
  - Such an information may greatly improve patients' treatment
- A simple "question" reveals personal information: The hospital B learns that Alice and Bob are being hospitalized in the hospital A.

# The notion of Private Set Intersection (PSI)




- PSI: Well-known privacy enhancing techniques for finding common entries between two different datasets
  - Only the common entries become known
  - A specific case of secure two-party computation
- Issues to be taken into account:
  - Each individual should be identified through the same unique identifier in each of the hospitals.
  - How the accuracy/authenticity of the input provided by each party is ensured?

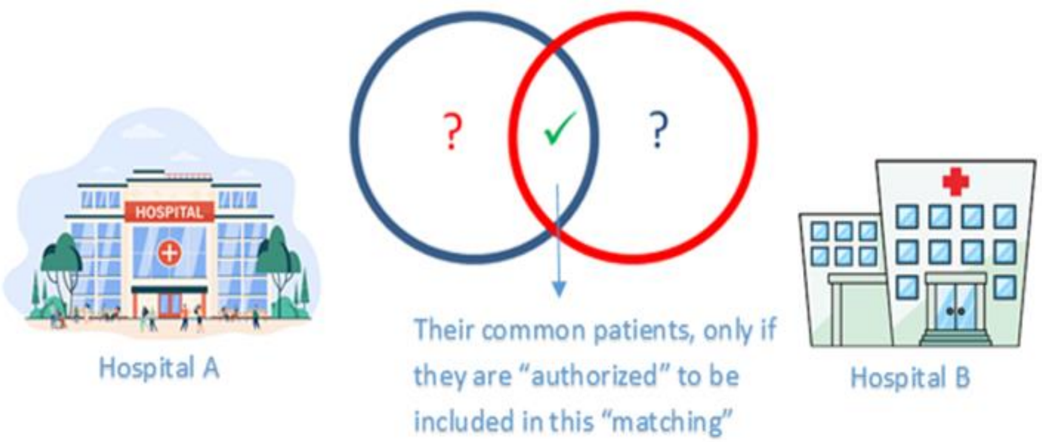
# The notion of Private Set Intersection (PSI) (*Cont.)*




- Especially in the health sector, more challenges occur even if PSI is to be applied:
  - There may be additional (legal) requirements
  - For example, the two hospitals may be allowed to share information about Alice only if both have her consent

- For all these reasons, even employing classical PSI techniques for data sharing within the health sector may not be (always) enough….

# Authorised Private Set Intersection (APSI)




- APSI: Each element in the set must be first authorized for sharing by a mutually trusted authority
  - It somehow resembles the Certification Authorities in Public Key Infrastructures.
    - For example, if user's consent is required, this authority verifies that only the data of those users provided their consent will be included in the datasets and will feed the PSI protocol.

# Summarizing…

- Data sharing in health sector introduces several domain-specific challenges from a data protection point of view.
  - Only some indicative scenarios have been given
- The state-of-the-art though, towards ensuring data protection by design, may be more powerful than we think….
  - To be appropriately considered on a risk-based approach, taking into account all the factors
- These techniques are also applicable in other domains

# …and concluding

- Efforts should be put on promoting further research, not only from an academic but also from a practical point of view
  - Can such approaches be post-quantum secure? How to deal with the "store-now-decrypt-later" attacks?
    - Recall that the "time life" for health data is big….
  - What about sharing data in the context of machine learning (ML) techniques for creating proper statistical models from medical data?
- Developing such "advanced" approaches into standards?

# Some references

**ENISA Reports**


- ENISA, "Data protection engineering", 2022
- ENISA, "Deploying pseudonymisation techniques: the case of health sector", 2022
- ENISA, "Data Pseudonymisation: Advanced Techniques and Use Cases", 2021.


**Others**


- F. Wang, J. Mickens, N. Zeldovich, V. Vaikuntanathan, "Sieve: Cryptographically Enforced Access Control for User Data in Untrusted Clouds", USENIX Symposium on Networked Systems Design and Implementation, 2016.
- B. E. van Gastel, B. Jacobs, J. Popma, "Data Protection Using Polymorphic Pseudonymisation in a Large-Scale Parkinson's Disease Study", Journal of Parkinson's Disease, 2021
- K. Limniotis, "Cryptography as the means to protect fundamental human rights", Cryptography, 2021.

# Thank you for your attention!

Dr. Konstantinos Limniotis

ICT Auditor, Hellenic Data Protection Authority

Adjunct Faculty Member, Open University of Cyprus

E-mail: klimniotis at dpa.gr