Το έργο χρηματοδοτήθηκε από το
Πρόγραμμα Δικαιώματα,
Ισότητα και Ιθαγένεια 2014-2020
της Ευρωπαϊκής Ένωσης

Pseudonymization

Facilitating GDPR compliance for SMEs and promoting Data Protection by Design in ICT products and services (www.bydesign-project.eu)

ΑΡΧΗ ΠΡΟΣΤΑΣΙΑΣ ΔΕΔΟΜΕΝΩΝ

ΚΕΝΤΡΟ ΕΡΕΥΝΩΝ ΠΑΝΕΠΙΣΤΗΜΙΟΥ ΠΕΙΡΑΙΩΣ

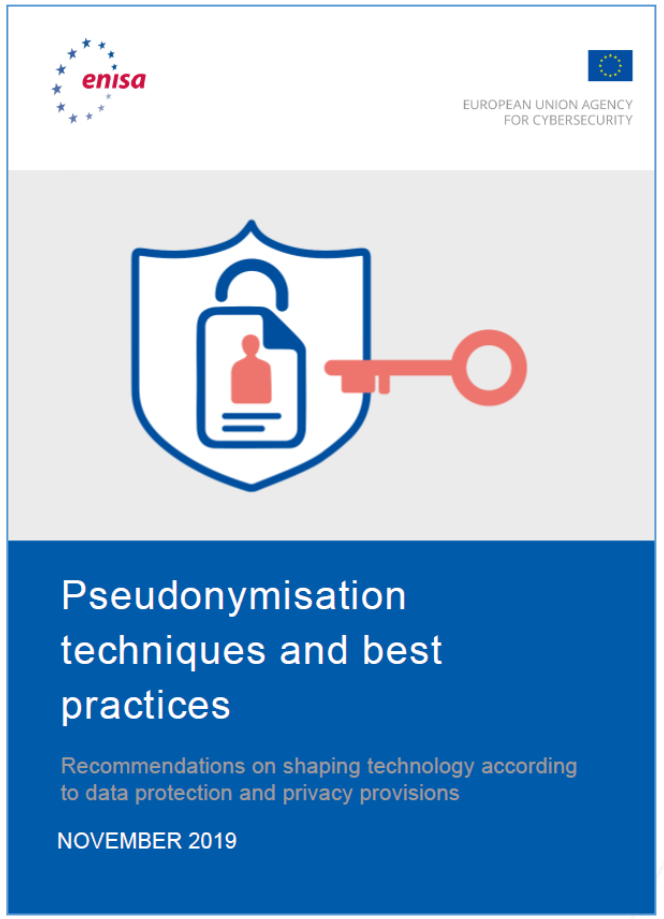
ICT abovo Information & Communication Technologies

This presentation has been based on material provided by Dr. K. Limniotis (HDPA)

Το έργο χρηματοδοτήθηκε από το Πρόγραμμα Δικαιώματα, Ισότητα και Ιθαγένεια 2014-2020 της Ευρωπαϊκής Ένωσης

ENISA Report: Pseudonymisation techniques and best practices

- Result of ENISA Project (March – October 2019)
  - **Editors**: Athena Bourka (ENISA)
    Prokopios Drogkaris (ENISA)
    Ioannis Agrafiotis (ENISA)
  - **Contributors**: Meiko Jensen (Kiel University)
    Cedric Lauradoux (INRIA)
    Konstantinos Limniotis (HDPA)
- Continuation of previous ENISA report
  - "**An overview on data pseudonymisation**", 2018



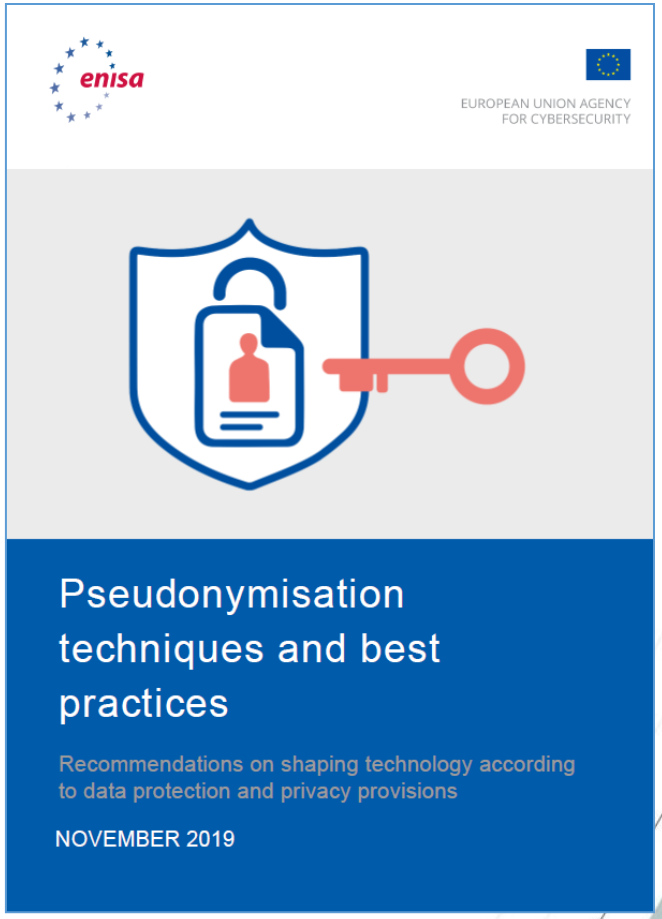
Το έργο χρηματοδοτήθηκε από το Πρόγραμμα Δικαιώματα, Ισότητα και Ιθαγένεια 2014-2020 της Ευρωπαϊκής Ένωσης

ENISA Report: Pseudonymisation techniques and best practices

- Focus on Techniques and Best Practices in Real-World Application Scenarios

- Terminology
- Scenarios
- Adversary Models
- Techniques
- Application Scenarios
  - IP Address pseudonymization
  - E-Mail Address pseudonymization
  - Pseudonymization in practice (discussion of complex cases)

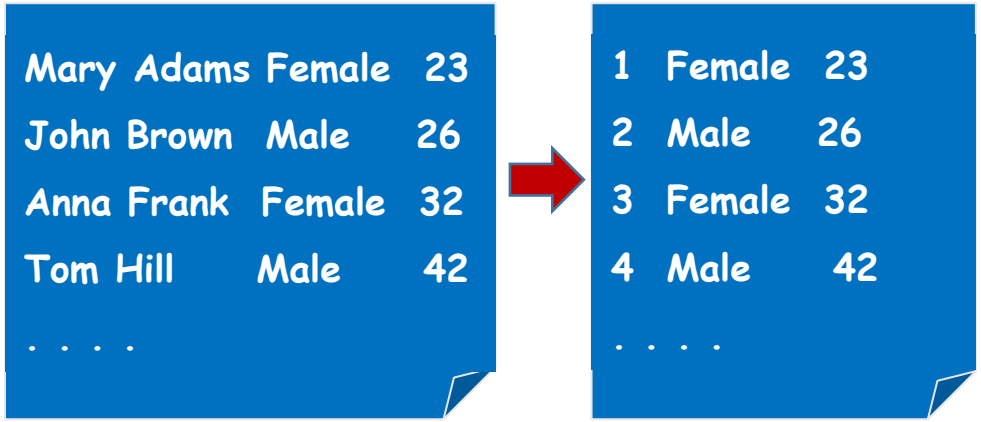
In relation with the desired «goals» of the pseudonymisation



General pseudonymisation goals

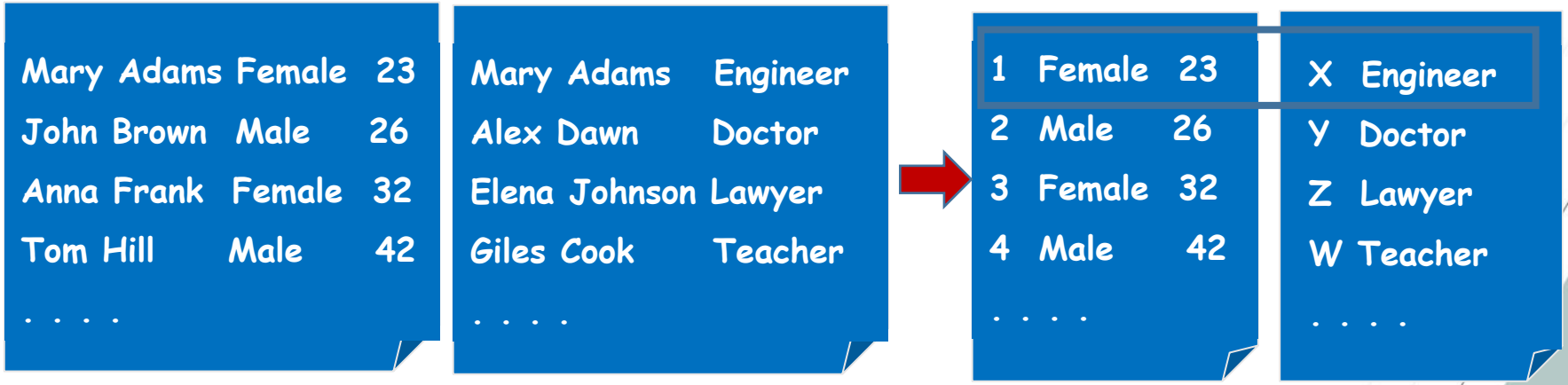
Το έργο χρηματοδοτήθηκε από το Πρόγραμμα Δικαιώματα, Ισότητα και Ιθαγένεια 2014-2020 της Ευρωπαϊκής Ένωσης

1. Hiding identities (related to confidentiality)



- Both goals are actually also related to the data minimization principle
- Be careful with the “confidentiality”: The pseudonymised data are not encrypted data (see next)
- Note that, in some cases, pseudonyms need to “carry” some information (i.e. increasing usability – see next), despite the fact that the identities should remain hidden

2. Unlinkability



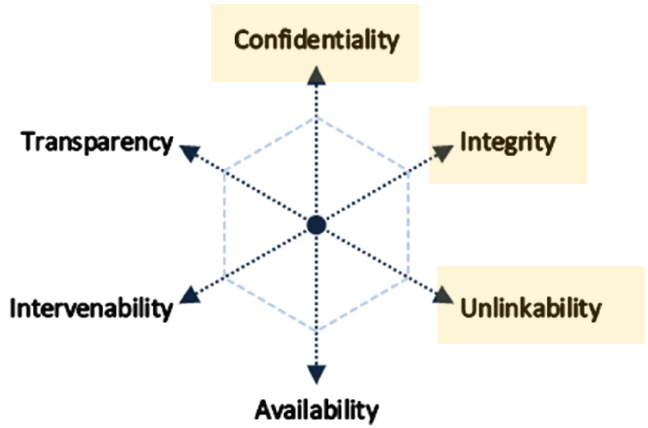
Το έργο χρηματοδοτήθηκε από το Πρόγραμμα Δικαιώματα, Ισότητα και Ιθαγένεια 2014-2020 της Ευρωπαϊκής Ένωσης

General pseudonymisation goals (*Cont.)*

3. Verification of the identity (related to integrity)

Mary Adams	→	EA2C1C52424CA8B645A3BD250350421F
Mary Edams	→	0A0FF77F292CA7B2C9B643548D8AB436
Mary Adams	→	EA2C1C52424CA8B645A3BD250350421F

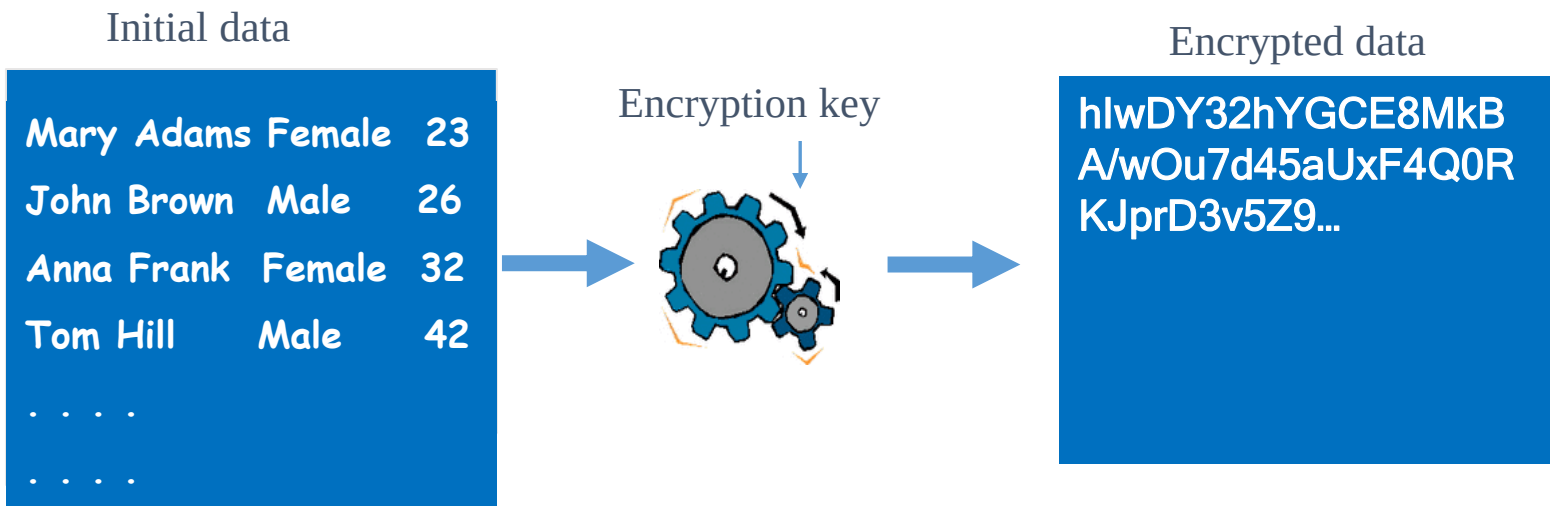
Summarizing: Pseudonymisation in relation to general data protection goals



M.Hansen, M. Jensen, and M. Rost, "Protection goals for privacy engineering", 2015

Pseudonymisation ≠ Encryption

Το έργο χρηματοδοτήθηκε από το Πρόγραμμα Δικαιώματα, Ισότητα και Ιθαγένεια 2014-2020 της Ευρωπαϊκής Ένωσης



- Encrypted data are unintelligible to anyone not having the decryption key (which inverses the encryption)
  - Not even statistical analysis can be performed on encrypted data
    - In general this is not the case in pseudonymisation
- Hence, the difference between pseudonymisation and encryption is obvious
  - However, appropriate use of cryptography may give rise to “good” pseudonymisation techniques…
  - The secret key could coincide with the “additional information needed for re-identification”

Το έργο χρηματοδοτήθηκε από το Πρόγραμμα Δικαιώματα, Ισότητα και Ιθαγένεια 2014-2020 της Ευρωπαϊκής Ένωσης

Terminology - Roles

- What roles are involved in a classic pseudonymization scenario? How are they named?

  ➔ «***Pseudonymization Entity***», «***Adversary***»

- How do these roles relate to the roles of GDPR?

  ➔ «***Data controller***», «***Data subject***», "***Data processor***»

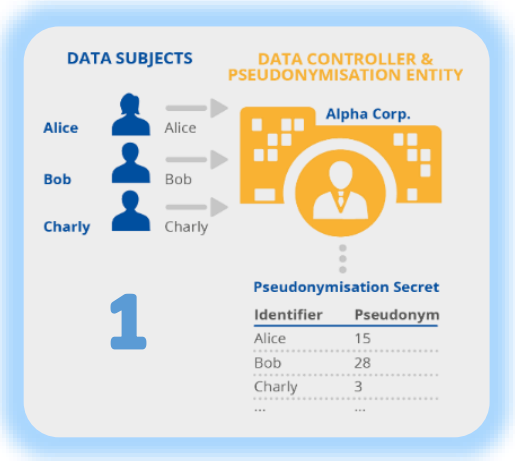
- **Encryption** is associated with a «secret key», but what is the «secret thing» of **pseudonymization**?
  - Related with the additional information needed for re-identification

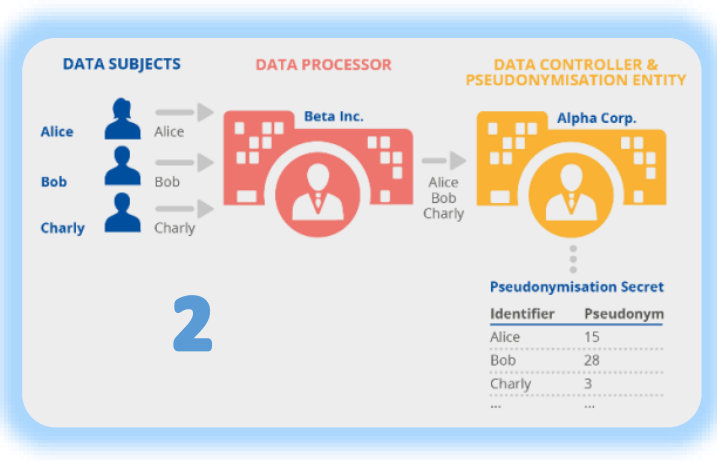
  ➔ «***Pseudonymization secret***»

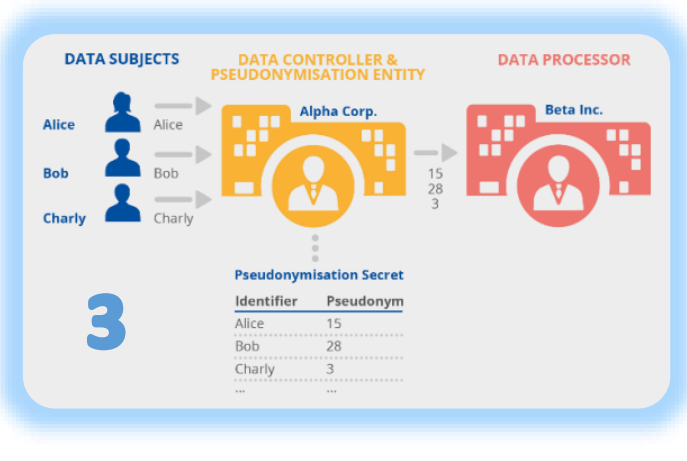
Scenarios

Το έργο χρηματοδοτήθηκε από το Πρόγραμμα Δικαιώματα, Ισότητα και Ιθαγένεια 2014-2020 της Ευρωπαϊκής Ένωσης

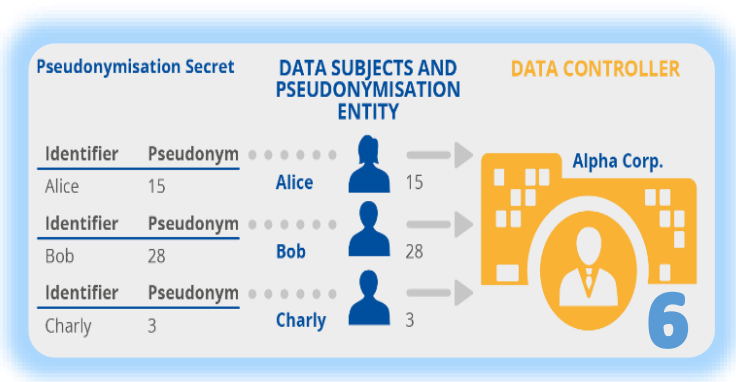
Scenarios 1-3: Data controller coincides with the pseudonymisation entity – i.e. the entity that actually performs pseudonymisation

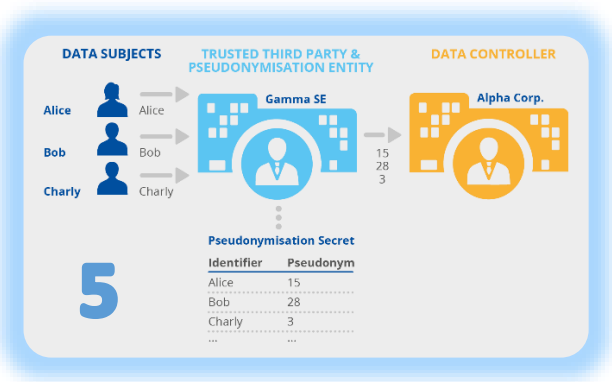


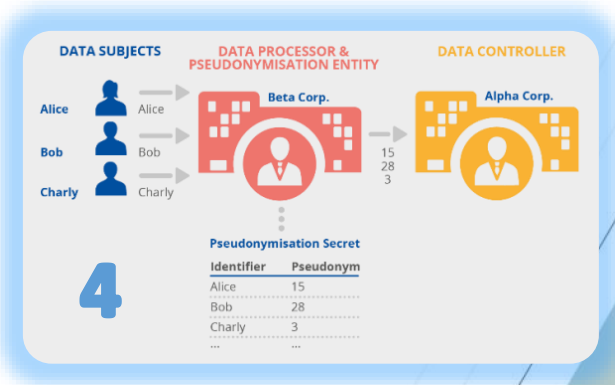




Scenarios 4-6: Data controller does **not** coincide with the pseudonymisation entity

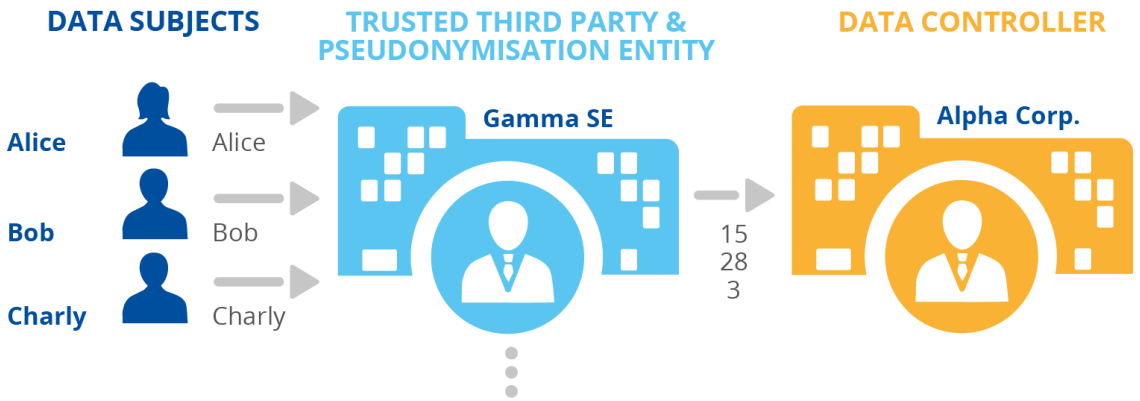






Το έργο χρηματοδοτήθηκε από το Πρόγραμμα Δικαιώματα, Ισότητα και Ιθαγένεια 2014-2020 της Ευρωπαϊκής Ένωσης

«Special» scenarios: Scenario 5



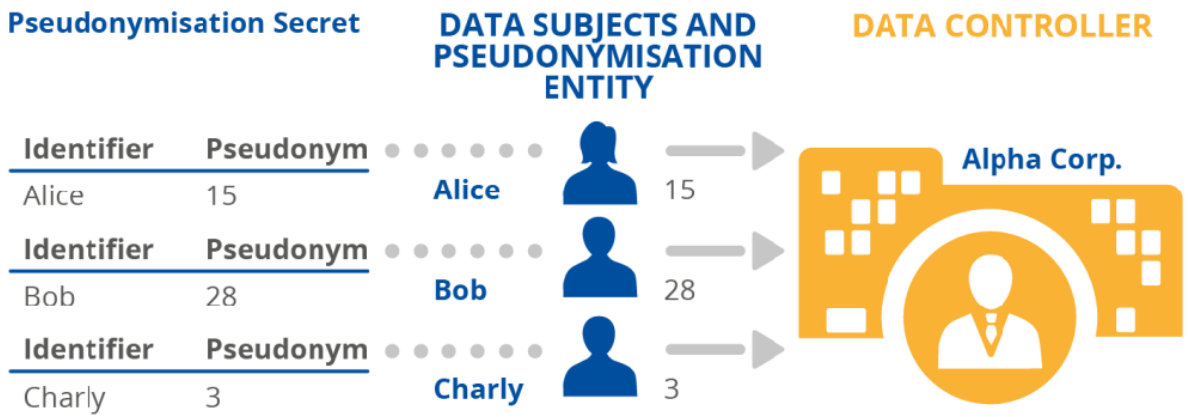
Trusted Third Party is always out of control of data controller!

Scenario 5

- ***Preceeding Trusted Third Party***
- ***Trusted Third Party = Pseudonymization Entity***
- ***Joint controllers??***

Το έργο χρηματοδοτήθηκε από το Πρόγραμμα Δικαιώματα, Ισότητα και Ιθαγένεια 2014-2020 της Ευρωπαϊκής Ένωσης

«Special» scenarios: Scenario 6



Do not confuse with the case that the data subjects simply choose their pseudonyms

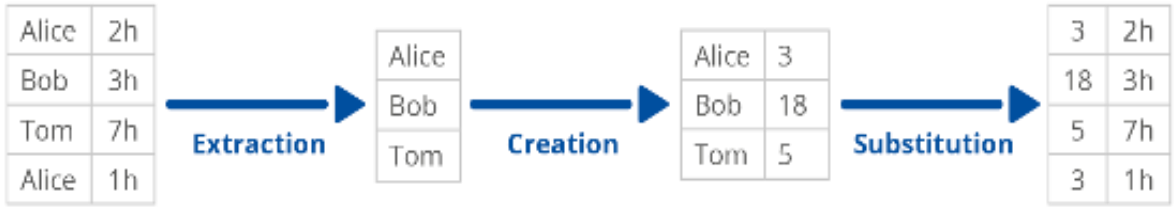
Scenario 6

- *The whole process is «governed" by the data controller*
- *Useful in cases that the data controller <u>should not know the original identities</u>*
- *E.g.: The ID of an electronic ticket for public transportation, uniquely constructed by a owner's passphrase (only the owner of the ticket knows the passphrase)*

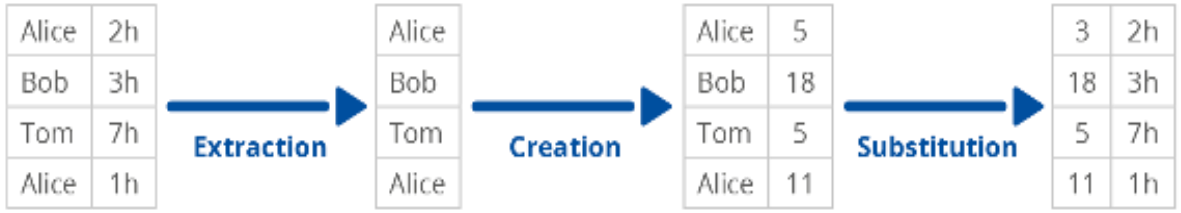
Το έργο χρηματοδοτήθηκε από το Πρόγραμμα Δικαιώματα, Ισότητα και Ιθαγένεια 2014-2020 της Ευρωπαϊκής Ένωσης

Pseudonymisation policies

1. Deterministic pseudonymisation



2. Randomized pseudonymisation



- The desired purpose of the pseudonymisation actually determines the policy that is preferable
- Deterministic pseudonymisation allows “tracking” of an individual within a database (more usability but also, probably, more data protection risks)

Pseudonymisation techniques

Το έργο χρηματοδοτήθηκε από το Πρόγραμμα Δικαιώματα, Ισότητα και Ιθαγένεια 2014-2020 της Ευρωπαϊκής Ένωσης

1. Counter / Random Number Generator (RNG)

"Hiding" everything

E-mail address	Pseudonym (Random number generator)	Pseudonym (counter generator)
alice@abc.eu	328	10
bob@wxyz.com	105	11
eve@abc.eu	209	12
john@qed.edu	83	13
alice@wxyz.com	512	14
mary@clm.eu	289	15

Keeping information on domains

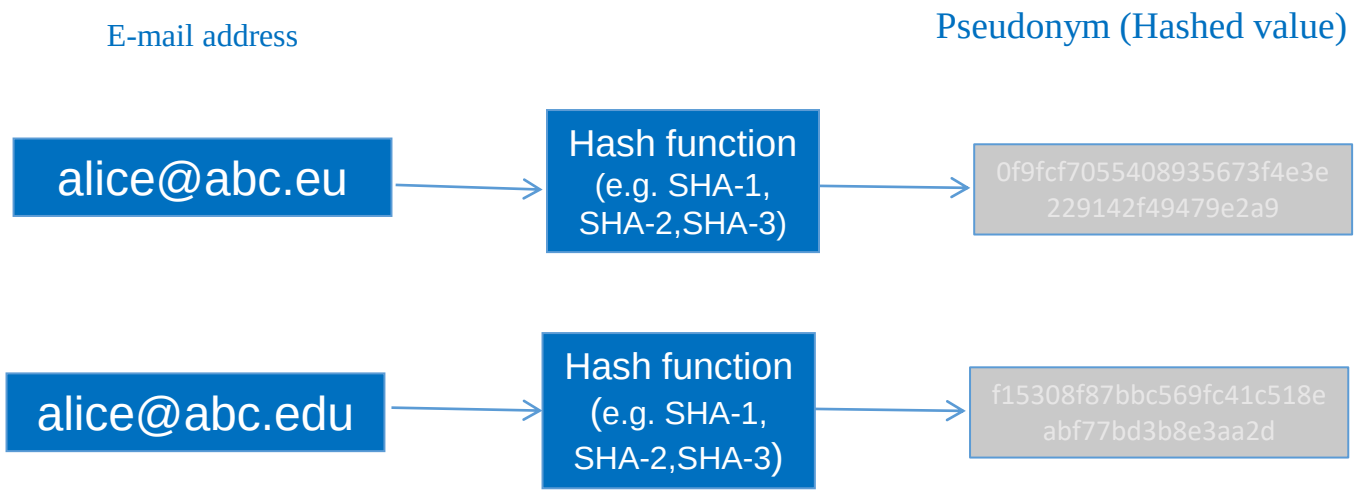
E-mail address	Pseudonym (Random number generator)	Pseudonym (counter generator)
alice@abc.eu	328@abc.eu	10@abc.eu
bob@wxyz.com	105@wxyz.com	11@wxyz.com
eve@abc.eu	209@abc.eu	12@abc.eu
john@qed.edu	83@qed.edu	13@qed.edu
alice@wxyz.com	512@wxyz.com	14@wxyz.com
mary@clm.eu	289@clm.eu	15@clm.eu

- Pseudonymisation secret = Mapping table
- Simplicity
- Scalability issues
  - Especially in deterministic pseudonymisation
- The counter-based pseudonyms may generally allow for some information extraction and/or prediction
  - (e.g. consider consecutive University students addresses, stud790@universityA.edu, stud791@universityA.edu etc.)

Pseudonymisation techniques

Το έργο χρηματοδοτήθηκε από το Πρόγραμμα Δικαιώματα, Ισότητα και Ιθαγένεια 2014-2020 της Ευρωπαϊκής Ένωσης

2. Cryptographic hash function

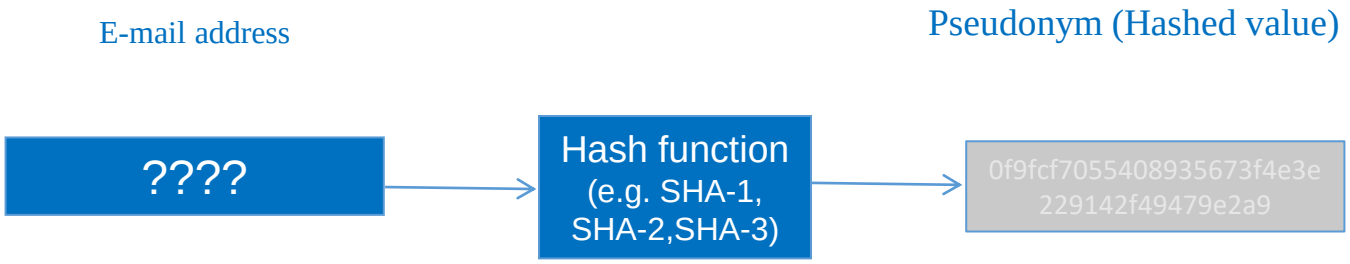


People believe that hashing is a nice pseudonymisation technique. **But…**

Pseudonymisation techniques

Το έργο χρηματοδοτήθηκε από το Πρόγραμμα Δικαιώματα, Ισότητα και Ιθαγένεια 2014-2020 της Ευρωπαϊκής Ένωσης

2. Cryptographic hash function *(Cont.)*



The adversary can easily verify whether any of the pseudonyms in the pseudonymised list corresponds to alice@abc.eu

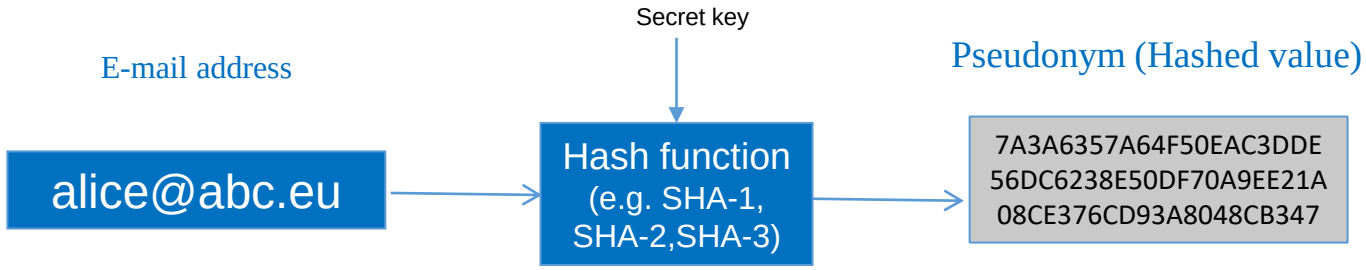
- Simply computes the hashed value of alice@abc.eu and checks…

- Actually, in such a scenario there is no pseudonymisation secret...
  - The only “secret” is the input domain
  - The size and the «predictability» of the input domain highly affects the level of protection (identity hiding) that a hash function provides as a pseudonymisation technique

Pseudonymisation techniques

Το έργο χρηματοδοτήθηκε από το Πρόγραμμα Δικαιώματα, Ισότητα και Ιθαγένεια 2014-2020 της Ευρωπαϊκής Ένωσης

3. Cryptographic hash function with key (Message Authentication Code – MAC)

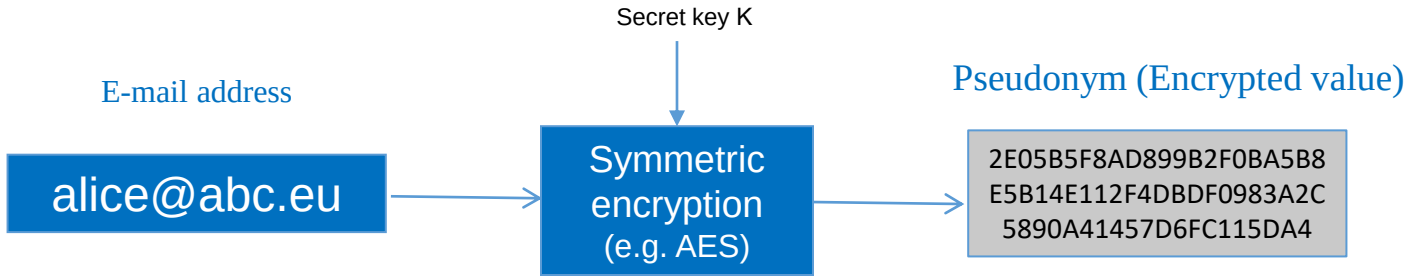


- Pseudonymisation secret = Secret key
- Deterministic or randomised pseudonymisation, based on whether the secret key is fixed or not
- High protection on «hiding» the initial identifier (once the key remains secret)
- High scalability
- But.. restrictions even for the pseudonymisation entity
  - Knowledge of the pseudonym and the pseudonymisation secret does not allow direct estimation of the initial identifier
    - However, given an identifier, it can be easily checked which is its corresponing pseudonym

Pseudonymisation techniques

Το έργο χρηματοδοτήθηκε από το Πρόγραμμα Δικαιώματα, Ισότητα και Ιθαγένεια 2014-2020 της Ευρωπαϊκής Ένωσης

4. Encryption – the deterministic case

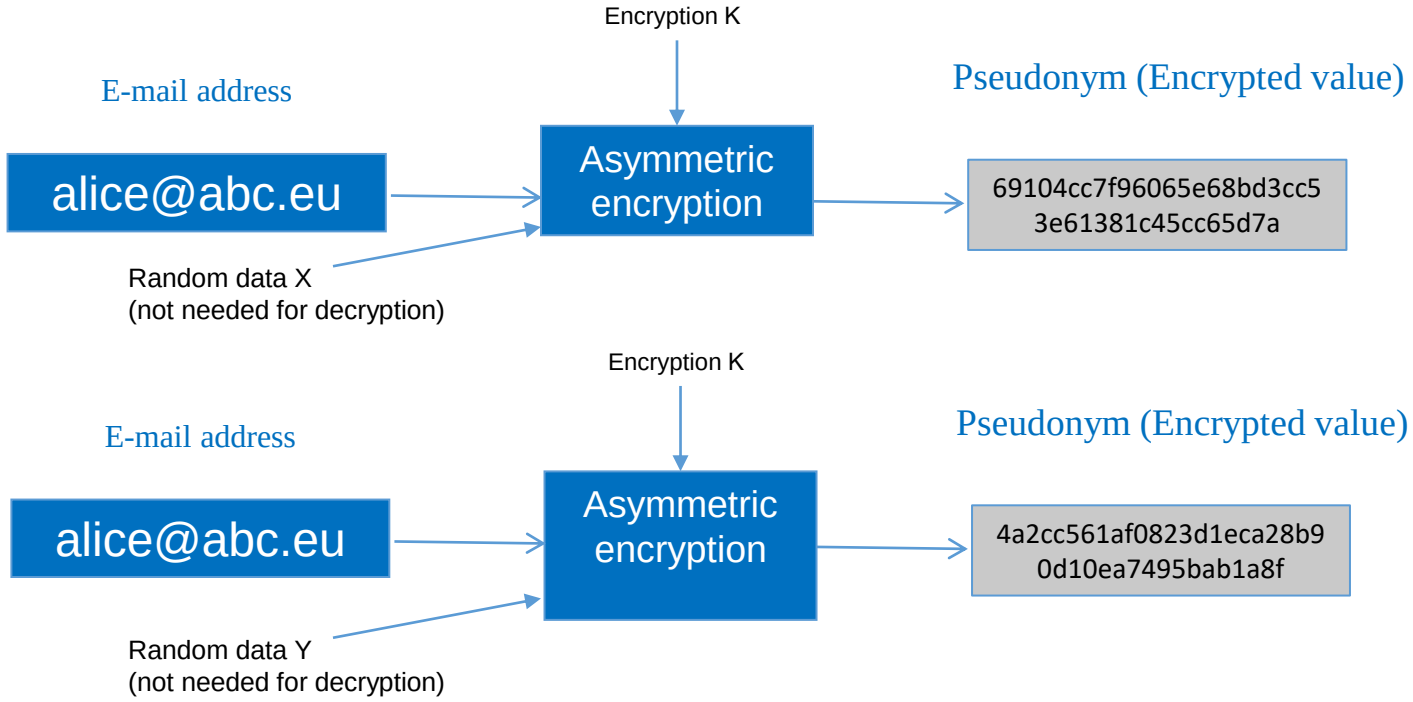


- Pseudonymisation secret = Secret key (the same for decryption)
- Deterministic pseudonymisation, for fixed secret key
- High protection on «hiding» the initial identifier (once the key remains secret)
- High scalability
- No restrictions for the pseudonymisation entity
  - Knowledge of the pseudonym and the pseudonymisation secret allows direct estimation of the initial identifier

Pseudonymisation techniques

Το έργο χρηματοδοτήθηκε από το Πρόγραμμα Δικαιώματα, Ισότητα και Ιθαγένεια 2014-2020 της Ευρωπαϊκής Ένωσης

4. Encryption – the probabilistic case



- Pseudonymisation secret = Decryption key (different from encryption key)
- Randomised pseudonymisation
- Other pseudonymisation benefits similar to deterministic encryption are also present