ΑΡΙΣΤΟΤΕΛΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΟΝΙΚΗΣ

Τεχνικές και Διαδικασίες Ψευδωνυμοποίησης βάσει του κανονισμού GDPR

Cybersecurity Seminar by Google

Αναφορά 2ης Εργασίας

Νίκος Τουλκερίδης

Ιανουάριος 2026

Περιεχόμενα

1 Ταξινόμηση Απλών Προσωπικών Δεδομένων

Με βάση τα δεδομένα της εκφώνησης και τους ορισμούς του GDPR (Διαφάνεια 9), η ταξινόμηση γίνεται ως εξής:

1.1 Άμεσα Αναγνωριστικά (Direct Identifiers)

Πρόκειται για στοιχεία που ταυτοποιούν μοναδικά και άμεσα το φυσικό πρόσωπο:

- **Ονοματεπώνυμο ασθενούς:** Αποτελεί το βασικό στοιχείο άμεσης αναγνώρισης.
- **ΑΜΚΑ:** Είναι μοναδικός αριθμός ταυτοποίησης για κάθε πολίτη.
- **Τηλέφωνο:** Είναι προσωπικό στοιχείο που συνδέεται άμεσα με τον κάτοχο.

1.2 Έμμεσα Αναγνωριστικά (Indirect Identifiers)

Πρόκειται για στοιχεία που από μόνα τους δεν αρκούν πάντα για μοναδική ταυτοποίηση, αλλά συνδυαστικά οδηγούν σε αυτή:

- **Διεύθυνση κατοικίας:** Θεωρείται έμμεσο αναγνωριστικό καθώς μπορεί να αφορά περισσότερα από ένα άτομα (π.χ. μέλη οικογένειας), ωστόσο περιορίζει σημαντικά το εύρος αναζήτησης.

2 Ευαίσθητα Προσωπικά Δεδομένα (Special Categories of Personal Data)

Σύμφωνα με το Άρθρο 9 του GDPR και τη διαφάνεια 9 της διάλεξης, τα παρακάτω δεδομένα κατατάσσονται στις “Ειδικές Κατηγορίες” (ευαίσθητα) καθώς αφορούν την υγεία του υποκειμένου:

- **Αποτελέσματα εξετάσεων (βιοχημικές & αιματολογικές):** Αποκαλύπτουν βιολογικά δεδομένα και την κατάσταση υγείας.
- **Ιστορικό θεραπειών:** Περιλαμβάνει πληροφορίες για την ιατρική περίθαλψη που έχει λάβει ο ασθενής.
- **Κλινική διάγνωση:** Αποτελεί ιατρικό συμπέρασμα για την κατάσταση της υγείας.

Σημείωση: Βάσει της διαφάνειας 26, επισημαίνεται πως σε περιβάλλον υγείας, ακόμη και απλά δεδομένα (όπως ΑΜΚΑ ή Όνομα) μπορούν να θεωρηθούν ευαίσθητα ως τμήμα του ιατρικού φακέλου, καθώς η συσχέτισή τους αποκαλύπτει την ιδιότητα του ασθενούς.

3 Πρόταση Μεθόδου Ψευδωνυμοποίησης (Βάσει Άρθρου 4 GDPR)

Για την αντικατάσταση των πραγματικών στοιχείων με τεχνητούς κωδικούς της μορφής “PAT001”, επιλέγεται η τεχνική του **Μετρητή (Counter)** από τον πίνακα της διαφάνειας 48.

3.1 Προτεινόμενη Τεχνική: Μετρητής (Counter)

Η τεχνική αυτή χρησιμοποιεί έναν μονότονο μετρητή που αυξάνεται κάθε φορά που καταχωρείται ένας νέος ασθενής (π.χ. PAT001, PAT002), δημιουργώντας ένα μοναδικό ψευδώνυμο που δεν έχει καμία λογική σχέση με τα αρχικά δεδομένα.

3.2 Διαδικασία Υλοποίησης

1. **Δημιουργία Πίνακα Αντιστοίχισης (Mapping Table):** Κατασκευάζεται ένας πίνακας που συνδέει τα Άμεσα Αναγνωριστικά (Ονοματεπώνυμο, ΑΜΚΑ, Τηλέφωνο, Διεύθυνση) με το Ψευδώνυμο (PATxxx).

2. **Αντικατάσταση & Διαχωρισμός:**

   - Στο αρχείο που θα σταλεί στο Πανεπιστήμιο, αφαιρούνται πλήρως τα πεδία: Ονοματεπώνυμο, ΑΜΚΑ, Τηλέφωνο και Διεύθυνση.
   - Τα πεδία αυτά αντικαθίστανται από το Ψευδώνυμο (π.χ. PAT001).
   - Τα Ιατρικά Δεδομένα (Εξετάσεις, Ιστορικό, Διάγνωση) παραμένουν, αλλά πλέον συνδέονται μόνο με το PAT001.

Αποτέλεσμα: Τα δεδομένα δεν μπορούν πλέον να αποδοθούν στο συγκεκριμένο φυσικό πρόσωπο χωρίς τη χρήση του Πίνακα Αντιστοίχισης.

4 Ασφαλής Αποθήκευση Πίνακα Αντιστοίχισης (Βάσει Άρθρου 32 GDPR)

Σύμφωνα με το Άρθρο 32 («Ασφάλεια της επεξεργασίας») και τον ορισμό της ψευδωνυμοποίησης στο Άρθρο 4, ο Πίνακας Αντιστοίχισης αποτελεί το «κλειδί» της αποκωδικοποίησης και απαιτεί τα εξής μέτρα ασφαλείας:

4.1 Λειτουργικός Διαχωρισμός (Separation)

Οι «συμπληρωματικές πληροφορίες» (δηλ. ο πίνακας) πρέπει να διατηρούνται χωριστά από τα ψευδωνυμοποιημένα δεδομένα. Ο πίνακας παραμένει στο εσωτερικό δίκτυο του Νοσοκομείου και δεν αποθηκεύεται ποτέ στον ίδιο φάκελο ή βάση δεδομένων με τα δεδομένα που θα σταλούν στο Πανεπιστήμιο.

4.2 Κρυπτογράφηση (Encryption)

Ο πίνακας πρέπει να αποθηκεύεται κρυπτογραφημένος (encryption at rest). Αυτό διασφαλίζει ότι ακόμη και αν κλαπεί το μέσο αποθήκευσης, η σύνδεση κωδικού-ασθενούς δεν θα είναι δυνατή χωρίς το κλειδί αποκρυπτογράφησης.

4.3 Περιορισμός Πρόσβασης (Access Control)

Η πρόσβαση στον πίνακα πρέπει να περιορίζεται αποκλειστικά σε εξουσιοδοτημένο προσωπικό του Νοσοκομείου (π.χ. υπεύθυνοι IT/ασφαλείας). Οι ερευνητές του Πανεπιστημίου δεν πρέπει να έχουν καμία πρόσβαση σε αυτόν.

5 Διαδικασία, Ρόλοι και Αξιολόγηση Κινδύνου

5.1 Βήματα Διαδικασίας (Workflow)

1. **Εξαγωγή:** Το τμήμα Πληροφορικής (IT) εξάγει τα επιλεγμένα δεδομένα από τη βάση του Νοσοκομείου.
2. **Διαχωρισμός & Ψευδωνυμοποίηση:** Τα Άμεσα Αναγνωριστικά αφαιρούνται και αντικαθίστανται από τους κωδικούς (PATxxx) βάσει του Πίνακα Αντιστοίχισης.
3. **Έλεγχος:** Επιβεβαιώνεται ότι δεν έχουν ξεμείνει αναγνωριστικά (π.χ. κάποιο όνομα σε πεδίο σχολίων).
4. **Διαβίβαση:** Το αρχείο που περιέχει μόνο τους κωδικούς και τα ιατρικά δεδομένα αποστέλλεται στο Πανεπιστήμιο μέσω ασφαλούς καναλιού.

5.2 Ρόλος DPO και Εμπλεκόμενοι

- **Προσωπικό IT/Ασφαλείας:** Είναι οι μόνοι που εκτελούν την τεχνική διαδικασία και έχουν πρόσβαση στον Πίνακα Αντιστοίχισης (“need-to-know” basis).
- **DPO (Υπεύθυνος Προστασίας Δεδομένων):** Έχει εποπτικό και συμβουλευτικό ρόλο. Δεν εκτελεί ο ίδιος την ψευδωνυμοποίηση (για αποφυγή σύγκρουσης συμφερόντων), αλλά ελέγχει αν τηρείται η διαδικασία και αν έχει εκπονηθεί η Εκτίμηση Αντικτύπου (DPIA).

5.3 Μείωση Κινδύνου & Περιορισμοί

5.3.1 Γιατί μειώνει τον κίνδυνο

Η ψευδωνυμοποίηση “σπάει” την άμεση σύνδεση δεδομένων-προσώπου. Αν διαρρεύσει το αρχείο από το Πανεπιστήμιο, τα δεδομένα δεν μπορούν να αποδοθούν σε συγκεκριμένους ασθενείς χωρίς τον πίνακα.

5.3.2 Πότε δεν είναι επαρκής

Η ψευδωνυμοποίηση δεν εγγυάται πλήρη ανωνυμία. Υπάρχει κίνδυνος **"Επίθεσης Συσχέτισης" (Linkage Attack)**, όπως φαίνεται στη διαφάνεια 56. Αν το Πανεπιστήμιο έχει πρόσβαση σε άλλες πηγές (π.χ. δημοσιεύματα, social media), μπορεί να ταυτοποιήσει έναν ασθενή συνδυάζοντας έμμεσα αναγνωριστικά (π.χ. Σπάνια ασθένεια + Ηλικία + Περιοχή).

6 ΑΣΚΗΣΗ 2: Ψευδωνυμοποίηση Δεδομένων Πελατών

6.1 Κώδικας Υλοποίησης (Python)

Για την αντικατάσταση των άμεσων αναγνωριστικών (Όνομα, Email) με ψευδώνυμα της μορφής "USERx", χρησιμοποιήθηκε η τεχνική του **Μετρητή (Counter)**. Ακολουθεί ο κώδικας που εκτελέστηκε:

```python
# 1. Initial Data (Raw Data)
customers = [
    {"name": "Άννα", "surname": "Παπαδοπούλου",
     "email": "anna@example.com", "age": 28,
     "profession": "Ιατρός"},
    {"name": "Κώστας", "surname": "Νικολάου",
     "email": "kostas@example.com", "age": 35,
     "profession": "Μηχανικός Η/Υ"},
    {"name": "Ιωάννα", "surname": "Γεωργίου",
     "email": "ioanna@example.com", "age": 22,
     "profession": "Καθηγήτρια"}
]

# Lists to store the separated data
pseudonymized_data = []
mapping_table = []

# 2. Processing Loop (Technique: Counter)
# We use a simple counter (1, 2, 3...) to generate the User ID.
counter = 1

for person in customers:
    # Generate the Pseudonym
    user_id = f"USER{counter}"

    # A. Create the Mapping Table (Secret Key)
    # This stores the link between the ID and the Real Identity.
    mapping_table.append({
        "Pseudonym": user_id,
        "Real_Name": person["name"],
        "Real_Surname": person["surname"],
        "Real_Email": person["email"]
    })
```

```
34
35      # B. Create the Pseudonymized Data (For the Partner)
36      # This stores the ID and the data needed for analysis (Age,
            Profession).
37      # Identifiers (Name, Email) are REMOVED.
38      pseudonymized_data.append({
39          "Pseudonym": user_id,
40          "Age": person["age"],
41          "Profession": person["profession"]
42      })
43
44      counter += 1
45
46  # 3. Output the Results
47  print("--- TABLE 1: PSEUDONYMIZED DATA (Sent to External Partner)
        ---")
48  print(f"{'Pseudonym':<10} | {'Age':<5} | {'Profession'}")
49  print("-" * 40)
50  for row in pseudonymized_data:
51      print(f"{row['Pseudonym']:<10} | {row['Age']:<5} | {row['
            Profession']}")
52
53  print("\n" + "="*50 + "\n")
54
55  print("--- TABLE 2: MAPPING TABLE (Kept Securely internally) ---"
        )
56  print(f"{'Pseudonym':<10} | {'Real_Email':<20} | {'Full Name'}")
57  print("-" * 50)
58  for row in mapping_table:
59      print(f"{row['Pseudonym']:<10} | {row['Real_Email']:<20} | {
            row['Real_Name']} {row['Real_Surname']}")
```

Listing 1: Υλοποίηση Ψευδωνυμοποίησης με Python

6.2 Πίνακες Εξόδου

6.2.1 Πίνακας Ψευδωνυμοποιημένων Δεδομένων (Προς Διαβίβαση)

Ο παρακάτω πίνακας δεν περιέχει άμεσα αναγνωριστικά και μπορεί να σταλεί στον εξωτερικό συνεργάτη για τη στατιστική ανάλυση.

Pseudonym	**Age**	**Profession**
USER1	28	Ιατρός
USER2	35	Μηχανικός Η/Υ
USER3	22	Καθηγήτρια

Πίνακας 1: Ψευδωνυμοποιημένα Δεδομένα για Στατιστική Ανάλυση

6.2.2 Πίνακας Αντιστοίχισης (Mapping Table - Εσωτερική Χρήση)

Ο πίνακας αυτός συνδέει τα ψευδώνυμα με τα πραγματικά πρόσωπα.

Pseudonym	Real_Email	Full Name
USER1	anna@example.com	Άννα Παπαδοπούλου
USER2	kostas@example.com	Κώστας Νικολάου
USER3	ioanna@example.com	Ιωάννα Γεωργίου

Πίνακας 2: Πίνακας Αντιστοίχισης (Αποθηκεύεται Ξεχωριστά)

Σημείωση Ασφαλείας: Βάσει του Άρθρου 32 του GDPR, ο πίνακας αντιστοίχισης αποθηκεύεται σε ξεχωριστό, ασφαλές περιβάλλον (διαχωρισμός καθηκόντων) και προστατεύεται με κρυπτογράφηση.

6.3 Σχόλιο GDPR: Γιατί η ψευδωνυμοποίηση προστατεύει τα δεδομένα;

Η ψευδωνυμοποίηση αποτελεί κρίσιμο μέτρο ασφαλείας διότι **διαχωρίζει την πληροφορία ταυτοποίησης από τα δεδομένα ανάλυσης**.

- **Μείωση Κινδύνου:** Εάν ο πίνακας με τα ψευδωνυμοποιημένα δεδομένα διαρρεύσει ή υποκλαπεί από τον συνεργάτη, δεν μπορεί να συνδεθεί με φυσικά πρόσωπα χωρίς τον πίνακα αντιστοίχισης.
- **Προστασία Ιδιωτικότητας:** Επιτρέπει την επεξεργασία δεδομένων για στατιστικούς σκοπούς (όπως ζητάει η άσκηση) χωρίς να εκτίθεται η ταυτότητα των υποκειμένων, ικανοποιώντας την αρχή της ελαχιστοποίησης των δεδομένων (Data Minimisation).