ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΠΜΣ ΕΦΑΡΜΟΣΜΕΝΗΣ ΟΙΚΟΝΟΜΙΚΗΣ

Κατεύθυνση: Διοίκηση Δημόσιων Οργανισμών και Επιχειρήσεων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«Ανωνυμοποίηση δεδομένων. Μελέτη, χρήση και εγγυήσεις για δεδομένα δημοσίων οργανισμών και ιδιωτικών φορέων».

του **Τακτικού Βάιου**

Επιβλέπουσα Καθηγήτρια: Τσιλίκα Κυριακή

Αναπληρώτρια Καθηγήτρια Πανεπιστημίου Θεσσαλίας

Βόλος 2023

Υπεύθυνη Δήλωση

Δηλώνω υπεύθυνα ότι είμαι συγγραφέας αυτής της διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της, είναι πλήρως αναγνωρισμένη και αναφέρεται στη διπλωματική εργασία. Επίσης έχω αναφέρει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επίσης βεβαιώνω ότι αυτή η πτυχιακή εργασία προετοιμάστηκε από εμένα προσωπικά ειδικά για τις απαιτήσεις του προγράμματος μεταπτυχιακών σπουδών στην Εφαρμοσμένη Οικονομική του Τμήματος Οικονομικών Επιστημών του Πανεπιστημίου Θεσσαλίας

Ο δηλών

Τακτικός Βάιος

Βόλος, Μάιος 2023

Ευχαριστίες

Με την παρούσα διπλωματική εργασία ολοκληρώνεται το πρόγραμμα μεταπτυχιακών σπουδών «Εφαρμοσμένη Οικονομική» του πανεπιστημίου Θεσσαλίας, με κατεύθυνση την «Διοίκηση Δημοσίων Οργανισμών και Επιχειρήσεων». Ευχαριστώ όλους τους καθηγητές για τις γνώσεις που μου πρόσφεραν σε όλο αυτό το διάστημα.

Ευχαριστίες και μόνο οφείλω να εκφράσω σε όσους βοήθησαν στην εκπόνηση αυτής της εργασίας σε φίλους, συναδέλφους, προϊσταμένους τμημάτων και διευθυντές οργανισμών.

Θερμές ευχαριστίες οφείλω επίσης στην επιβλέπουσα καθηγήτριά μου κ. Τσιλίκα Κυριακή, για τη βοήθεια, την υποστήριξη, την υπομονή και την κατανόηση που υπέδειξε και λειτούργησε καταλυτικά στην ολοκλήρωση αυτής της προσπάθειας.

Η παρούσα εργασία θα ήταν αδύνατο να ολοκληρωθεί χωρίς την υποστήριξη της συζύγου μου Ελπίδας, που ανέλαβε την ψυχολογική και συναισθηματική στήριξή μου όλο αυτό το διάστημα και το μερίδιό μου στα οικογενειακά «βάρη».

Τέλος, αυτή η διπλωματική εργασία είναι αφιερωμένη στο γιο μου Νικηφόρο. Να στοχεύεις ψηλά γιε μου και να φέρεις πάντα νίκες.

Περίληψη

Η παρούσα μελέτη έχει ως σκοπό την γνωριμία και εξοικείωση με έννοιες και νομοθεσίες που αφορούν τα ευαίσθητα προσωπικά δεδομένα στα πλαίσια του Γενικού Κανονισμού για την Προστασία των Δεδομένων (ΓΚΠΔ), όπως αυτός εφαρμόζεται στον δημόσιο και ιδιωτικό τομέα. Επίσης θα παρουσιάσουμε τη διαδικασία ανωνυμοποίησης συνόλων δεδομένων, με το ελληνικής δημιουργίας ελεύθερο λογισμικό ανωνυμοποίησης Amnesia, το οποίο για την επίτευξη των στόχων του χρησιμοποιεί κανόνες γενίκευσης δεδομένων και αλγόριθμους k-ανωνυμίας. Θα αναφερθούμε στη λειτουργία και την εμπειρία χρήσης του, στα πλεονεκτήματα και στα μειονεκτήματά του και θα το συγκρίνουμε με ανταγωνιστές της παγκόσμιας αγοράς.

Τα παραδείγματα που θα αναλύσουμε αφορούν πραγματικά σενάρια συνόλων δεδομένων, τα οποία θα τα ανωνυμοποιήσουμε για να τα δημοσιεύσουμε με ασφάλεια, όχι μόνο για τα υποκείμενα αλλά και για τους οργανισμούς εντός των πλαισίων που θέτει ο ΓΚΠΔ, παρέχοντας βοήθεια στους μελλοντικούς χρήστες ώστε να εξοικειωθούν γρηγορότερα με όρους και έννοιες που σίγουρα θα χρειαστούν.

Για την επίτευξη του στόχου της παρούσης μελέτης έχουμε πραγματοποιήσει ποιοτική έρευνα με συνεντεύξεις στελεχών του δημόσιου και του ιδιωτικού τομέα αναζητώντας τη γνώση και την εμπειρία τους στην εφαρμογή του πλαισίου του ΓΚΠΔ, καθώς και τη δυνατότητα χρήσης τεχνικών ανωνυμοποίησης και προστασίας των ευαίσθητων προσωπικών δεδομένων που διαχειρίζονται.

Λέξεις – κλειδιά: ανωνυμοποίηση, λογισμικό ανωνυμοποίησης δεδομένων Amnesia, προσωπικά δεδομένα, ΓΚΠΔ, k-ανωνυμία

Abstract

The purpose of this study is to get to know and become familiar with concepts and legislation concerning sensitive personal data, within the framework of the General Data Protection Regulation (GDPR), as it is applied in the public and private sector. We will also present the process of anonymizing datasets with the Greek-created Amnesia anonymization software, which uses data generalization rules and k-anonymity algorithms to achieve its goals. We'll cover its functionality and user experience, its pros and cons, and compare it to global market competitors.

The examples that we will analyze concern real scenarios datasets, which we will anonymize for safe publication, not only for the subjects themselves but for the organizations also, within the frameworks set by the GDPR, in order to help future users to familiarize themselves more quickly with terms and concepts that they will definitely need.

To achieve the goal of this study, we have conducted qualitative research with interviews of public and private sector executives seeking their knowledge and experience in the application of the GDPR framework, as well as the possibility of using anonymization techniques and protection of the sensitive personal data they manage .

Keywords: anonymization, Amnesia anonymization tool, personal data, GDPR, k-anonymity

Περιεχόμενα

Περιεχόμενα εικόνων

Εισαγωγή

Δεν μπορεί να διαφωνήσει κανείς ότι διανύουμε μια εποχή όπου τα πάντα έχουν ψηφιακό πρόσωπο είτε αυτό ονομάζεται άτομο ως οντότητα στα κοινωνικά δίκτυα, είτε ως επιχείρηση, εταιρία ή οργανισμός με την δημιουργία ενός ψηφιακού προφίλ. Το διαδίκτυο αποτελεί μία ανεξάντλητη πηγή παντός είδους πληροφορίας που μας δίνει την δυνατότητα να πραγματοποιούμε ότι χρειαζόμαστε από την ασφάλεια του σπιτιού μας εξοικονομώντας χρόνο και χρήμα επενδύοντας σε άλλες εκφάνσεις της καθημερινότητάς μας. Με την αποδοχή αυτής της ευκολίας σε κάθε περίπτωση πρέπει να αναγνωρίσουμε και τους κινδύνους από μία ενδεχόμενη διαρροή προσωπικών στοιχείων με την μορφή ψηφιακών δεδομένων, ειδικά ύστερα από μία κακόβουλη επίθεση, καθιστώντας την εξεύρεση προστασίας ως μοναδική διαδικασία. (United Nations Development Group (2017); University of North Alambama, (accessed 2023)

Χαρακτηριστικό παράδειγμα των ανωτέρω είναι η επίκληση ενός άρθρου έρευνας του Security Magazine από το Πανεπιστήμιο της Βόρειας Αλαμπάμα στο οποίο αναφέρεται ότι από τα τέλη του 2019 υπήρξαν 7098 παραβιάσεις δεδομένων με πάνω από 15,1 τρισεκατομμύρια αρχεία δεδομένων πάρα πολύ γνωστών εταιριών και εφαρμογών. (University of North Alambama (accessed 2023) . Επίσης γνωστές είναι και οι περιπτώσεις του βραβείου Netflix και άλλων στις οποίες αναγνωρίσθηκαν οι πελάτες – χρήστες, ως φυσικά πρόσωπα παρόλες τις προσπάθειες τήρησης απορρήτου, όπως θα αναλύσουμε παρακάτω.

Επιπροσθέτως σύμφωνα με το σκεπτικό αρ. 101 της Ευρωπαϊκής Ένωσης για τον ΓΚΠΔ, λόγω της διόγκωσης των ροών προσωπικών δεδομένων, για την ανάπτυξη της διαδικτυακής συνεργασίας διεθνώς, κρίθηκε απαραίτητη η προστασίας κάθε δεδομένου προσωπικού χαρακτήρα. (ΓΚΠΔ (2016), σκεπτικό αρ. 101)

Όσον αφορά την ασφάλεια έχουμε τρεις κυρίαρχες τάσεις. Η πρώτη είναι αυτή κολοσσών όπως η Microsoft, Apple και Google, που επιθυμούν να συγκεντρώσουν πάνω τους, όλες τις δυνατές διασυνδέσεις του εκάστοτε χρήστη χρησιμοποιώντας τη λογική του ελέγχου αυθεντικοποίησης του χρήστη «δύο παραγόντων», όπου ο χρήστης θα χρησιμοποιεί μια επιπλέον συσκευή για την αναγνώρισή του (π.χ. smartphone). (Chandrakar & Om, 2015; Google AI Research (accessed 2023); Apple Differential Privacy (accessed 2023).;Harvard University Privacy Tools Project (accessed 2023).; Yin et al., 2020)

Η δεύτερη λύση έρχεται από τις ίδιες τις εταιρείες (και τις εταιρίες που φροντίζουν για την ασφάλεια αυτών), οι οποίες ενθαρρύνουν τους χρήστες στη δημιουργία και απομνημόνευση μοναδικών συνδυασμών πρόσβασης με συγκεκριμένες κατευθυντήριες γραμμές, όπως πχ με ελάχιστο αριθμό στοιχείων ή με συνδυασμό αριθμών και κεφαλαίων και πεζών γραμμάτων για την είσοδο στους ιστότοπους ή σε εφαρμογές (Yıldırım & Mackie, 2019), με όλες τις ανθρώπινες αδυναμίες σχετικά με τη δημιουργία, την απομνημόνευση και τη χρήση αυτών των κωδικών. (Julkunen & Ceder Molander, 2016)

Τι γίνεται όμως όταν χρειαζόμαστε να εξάγουμε δεδομένα για χρήση, μελέτη και έρευνα από την επιστημονική κοινότητα; Η Τρίτη λύση, αυτή που θα ερευνήσουμε στην παρούσα μελέτη, είναι η προστασία των ίδιων των δεδομένων μέσω λογισμικών που εφαρμόζουν αλγόριθμους ανωνυμοποίησης και ιεραρχίες γενίκευσης, ώστε να μπορούν μεν να μελετηθούν ή ακόμη και να διαμοιραστούν, χωρίς όμως να μπορούν να αναγνωστούν η να υπονομευτούν.

Η διαδικασία ανωνυμοποίησης δεδομένων παρουσιάζεται και αναλύεται μέσα από την επίδειξη και αξιολόγηση ενός λογισμικού ελληνικής δημιουργίας ανοιχτού κώδικα με την ονομασία Amnesia (Amnesia software v. 1.3.2 (04/2023))

Η μελέτη αυτή χωρίζεται σε επτά κεφάλαια:

Κεφάλαιο πρώτο όπου παρουσιάζονται και επεξηγούνται όλες οι βασικές έννοιες και οι ορισμοί,

Κεφάλαιο δεύτερο όπου περιγράφονται πραγματικά συμβάντα διαρροής προσωπικών δεδομένων,

Κεφάλαιο τρίτο όπου πραγματοποιούμε πρακτικά παραδείγματα ανωνυμοποίησης με παρουσίαση και ανάλυση του Amnesia, ενός λογισμικού ελληνικής δημιουργίας

Κεφάλαιο τέταρτο ο ανταγωνισμός και οι εναλλακτικές εφαρμογές στην παγκόσμια αγορά

Κεφάλαιο πέμπτο όπου καταδεικνύουμε τα πλεονεκτήματα και τα μειονεκτήματα του ελληνικού λογισμικού,

Κεφάλαιο έκτο όπου παρουσιάζουμε τα αποτελέσματα των συνεντεύξεων που πραγματοποιήσαμε πάνω στο αντικείμενο έρευνάς μας και

Κεφάλαιο έβδομο με τα σύνοψη και κλείσιμο αυτής της εργασίας

Σκοπός της Εργασίας – Προστιθέμενη Αξία

Ο σκοπός της παρούσης εργασίας δεν είναι η δημιουργία ενός οδηγού χρήσεως λογισμικών ανωνυμοποίησης. Η προστιθέμενη αξία αυτής της έρευνας είναι να βοηθήσει τους υπαλλήλους που εργάζονται στον δημόσιο ή στον ιδιωτικό τομέα, που διαχειρίζονται ή πρόκειται να διαχειριστούν προσωπικά δεδομένα, να κατανοήσουν έννοιες και νομοθεσίες σχετικές με την προστασία των προσωπικών δεδομένων. Να γνωρίσουν και να εξοικειωθούν με τα διάφορα λογισμικά που προσφέρουν ανωνυμοποίηση, μέσω πρακτικών παραδειγμάτων και εφαρμογών, ώστε μέσω της παρούσης διπλωματικής εργασίας να μπορέσουν να διαχειριστούν την ανωνυμοποίηση των προσωπικών δεδομένων της εταιρίας ή του οργανισμού που εργάζονται, με τρόπο νόμιμο και ασφαλή, τόσο για τα προσωπικά δεδομένα που διαχειρίζονται, όσο και για τους ίδιους και τις εταιρίες ή τους οργανισμούς που εκπροσωπούν. Επιπροσθέτως μέσω της παρουσίασης των λογισμικών ανωνυμοποίησης θα αναδείξουμε την τεχνολογική θέση στην οποία βρίσκεται η χώρας μας συγκριτικά με τον ανταγωνισμό και τις δυνατότητες βελτίωσής της.

ΚΕΦΑΛΑΙΟ 1

1. Βασικές έννοιες - Ορισμοί

1.1.Γενικός Κανονισμός Προστασίας Δεδομένων - ΓΚΠΔ

Σύμφωνα με το Γενικό Κανονισμό Προστασίας των δεδομένων ΓΚΠΔ – διεθνώς General Data Protection Regulation (GDPR) (2016) στο άρθρο 4 αναγράφονται ως ορισμοί:

«δεδομένα προσωπικού χαρακτήρα»:

κάθε πληροφορία που αφορά ταυτοποιημένο ή ταυτοποιήσιμο φυσικό πρόσωπο («υποκείμενο των δεδομένων»)· το ταυτοποιήσιμο φυσικό πρόσωπο είναι εκείνο του οποίου η ταυτότητα μπορεί να εξακριβωθεί, άμεσα ή έμμεσα, ιδίως μέσω αναφοράς σε αναγνωριστικό στοιχείο ταυτότητας, όπως όνομα, σε αριθμό ταυτότητας, σε δεδομένα θέσης, σε επιγραμμικό αναγνωριστικό ταυτότητας ή σε έναν ή περισσότερους παράγοντες που προσιδιάζουν στη σωματική, φυσιολογική, γενετική, ψυχολογική, οικονομική, πολιτιστική ή κοινωνική ταυτότητα του εν λόγω φυσικού προσώπου

«επεξεργασία»:

κάθε πράξη ή σειρά πράξεων που πραγματοποιείται με ή χωρίς τη χρήση αυτοματοποιημένων μέσων, σε δεδομένα προσωπικού χαρακτήρα ή σε σύνολα δεδομένων προσωπικού χαρακτήρα, όπως η συλλογή, η καταχώριση, η οργάνωση, η διάρθρωση, η αποθήκευση, η προσαρμογή ή η μεταβολή, η ανάκτηση, η αναζήτηση πληροφοριών, η χρήση, η κοινολόγηση με διαβίβαση, η διάδοση ή κάθε άλλη μορφή διάθεσης, η συσχέτιση ή ο συνδυασμός, ο περιορισμός, η διαγραφή ή η καταστροφή,

«περιορισμός της επεξεργασίας»:

η επισήμανση αποθηκευμένων δεδομένων προσωπικού χαρακτήρα με στόχο τον περιορισμό της επεξεργασίας τους στο μέλλον

«υπεύθυνος επεξεργασίας»:

το φυσικό ή νομικό πρόσωπο, η δημόσια αρχή, η υπηρεσία ή άλλος φορέας που, μόνα ή από κοινού με άλλα, καθορίζουν τους σκοπούς και τον τρόπο της επεξεργασίας δεδομένων προσωπικού χαρακτήρα· όταν οι σκοποί και ο τρόπος της επεξεργασίας αυτής καθορίζονται από το δίκαιο της Ένωσης ή το δίκαιο κράτους μέλους, ο υπεύθυνος επεξεργασίας ή τα ειδικά κριτήρια

για τον διορισμό του μπορούν να προβλέπονται από το δίκαιο της Ένωσης ή το δίκαιο κράτους μέλους,

«εκτελών την επεξεργασία»:

το φυσικό ή νομικό πρόσωπο, η δημόσια αρχή, η υπηρεσία ή άλλος φορέας που επεξεργάζεται δεδομένα προσωπικού χαρακτήρα για λογαριασμό του υπευθύνου της επεξεργασίας (ΓΚΠΔ – GDPR άρθρο 4)(*EUR-Lex - 32016R0679 - EN - EUR-Lex*; Ευρωπαϊκό Κοινοβούλιο, p. 42)

1.2.Ανωνυμοποίηση (Anonymization)

Σύμφωνα με το σκεπτικό του ΓΚΠΔ, ανωνυμοποίηση είναι η διαδικασία της ανωνυμίας των προσωπικών δεδομένων κατά τρόπο ώστε η ταυτότητα του υποκειμένου των δεδομένων να μην μπορεί ή να μην μπορεί πλέον να εξακριβωθεί. (Ευρωπαϊκό Κοινοβούλιο, p. 6)

Επιπροσθέτως με τους Meurers et al., η ανωνυμοποίηση δεδομένων είναι ένα σημαντικό δομικό στοιχείο για τη διασφάλιση του απορρήτου και προωθεί την επαναχρησιμοποίηση των δεδομένων.

Ωστόσο, ο μετασχηματισμός των δεδομένων με τρόπο που διατηρεί το απόρρητο των υποκειμένων, διατηρώντας παράλληλα υψηλό βαθμό ποιότητας δεδομένων, αποτελεί πρόκληση και ιδιαίτερα δύσκολη κατά την επεξεργασία σύνθετων συνόλων δεδομένων που περιέχουν μεγάλο αριθμό χαρακτηριστικών, με πρώτο βήμα την αφαίρεση όλων των χαρακτηριστικών που προσδιορίζουν άμεσα τα άτομα.(Meurers et al., 2021)

1.3.Ψευδωνυμοποίηση (Pseudonymization)

Η ψευδωνυμοποίηση είναι μια μέθοδος αποταυτοποίησης που αφαιρεί ή αντικαθιστά άμεσα αναγνωριστικά (ονόματα, ταυτότητες, αριθμούς τηλεφώνου κ.λπ.) από ένα ευαίσθητο σύνολο δεδομένων (αρχεία υγείας, ιατρικές συνταγές, οικονομικές πληροφορίες, διαδικτυακές έρευνες, αρχεία στο χώρο εργασίας κ.λπ.) (Amnesia software v. 1.3.2 (04/2023). Σύμφωνα με το Εθνικό Κέντρο Δημόσιας Διοίκησης και Αυτοδιοίκησης (ΕΚΔΔΑ), η ψευδωνυμοποίηση είναι η αντικατάσταση ενός ή περισσοτέρων αναγνωριστικών στοιχείων (identifiers), που επιτρέπουν την άμεση ή έμμεση ταυτοποίηση του ατόμου-υποκειμένου με ψευδώνυμα, καθώς επίσης και η προστασία και ο διαχωρισμός κάθε συμπληρωματικών πληροφοριών από τα τελικώς ψευδωνυμοποιημένα δεδομένα. (Σιουγλέ κα., 2022). Σύμφωνα με το ΓΚΠΔ:

«ψευδωνυμοποίηση»: η επεξεργασία δεδομένων προσωπικού χαρακτήρα κατά τρόπο ώστε τα δεδομένα να μην μπορούν πλέον να αποδοθούν σε συγκεκριμένο υποκείμενο των δεδομένων χωρίς τη χρήση συμπληρωματικών πληροφοριών, εφόσον οι εν λόγω συμπληρωματικές πληροφορίες διατηρούνται χωριστά και υπόκεινται σε τεχνικά και οργανωτικά μέτρα προκειμένου να διασφαλιστεί ότι δεν μπορούν να αποδοθούν σε ταυτοποιημένο ή ταυτοποιήσιμο φυσικό πρόσωπο. (Ευρωπαϊκό Κοινοβούλιο, p. 76)

Ο ΓΚΠΔ επιτρέπει την ανάλυση και επεξεργασία των προσωπικών δεδομένων με προϋπόθεση *την ύπαρξη κατάλληλων εγγυήσεων, που μπορεί να περιλαμβάνουν κρυπτογράφηση ή ψευδωνυμοποίηση.* (Ευρωπαϊκό Κοινοβούλιο, p. 46)

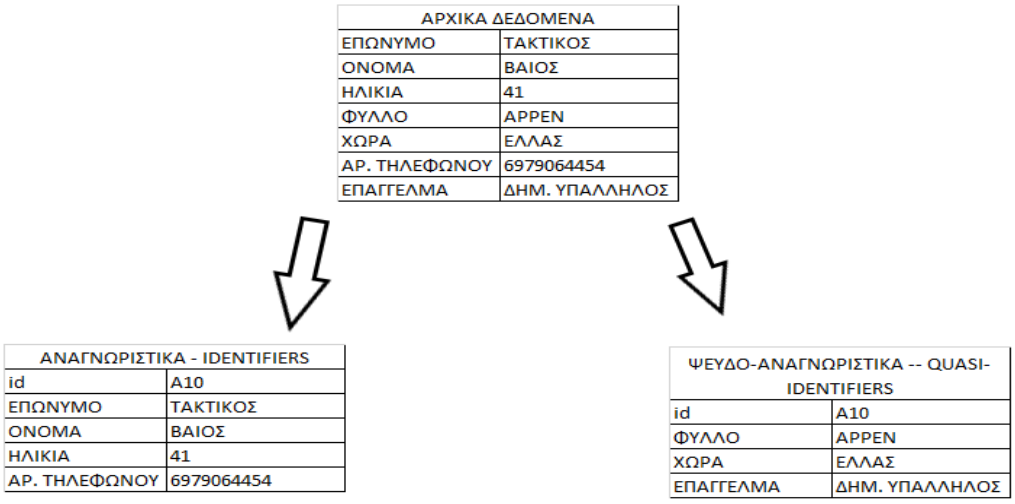
Επιπροσθέτως αναφέρεται πως: *ο υπεύθυνος επεξεργασίας εφαρμόζει αποτελεσματικά, τόσο κατά τη στιγμή του καθορισμού των μέσων επεξεργασίας όσο και κατά τη στιγμή της επεξεργασίας, κατάλληλα τεχνικά και οργανωτικά μέτρα, όπως η ψευδωνυμοποίηση, σχεδιασμένα για την εφαρμογή αρχών προστασίας των δεδομένων, όπως η ελαχιστοποίηση των δεδομένων, και την ενσωμάτωση των απαραίτητων εγγυήσεων στην επεξεργασία κατά τρόπο ώστε να πληρούνται οι απαιτήσεις του παρόντος κανονισμού και να προστατεύονται τα δικαιώματα των υποκειμένων των δεδομένων.* (Ευρωπαϊκό Κοινοβούλιο, p. 61)

Αναφορικά με την Ασφάλεια επεξεργασίας των δεδομένων αναφέρεται πως: *ο υπεύθυνος επεξεργασίας και ο εκτελών την επεξεργασία εφαρμόζουν κατάλληλα τεχνικά και οργανωτικά μέτρα προκειμένου να διασφαλίζεται το κατάλληλο επίπεδο ασφάλειας έναντι των κινδύνων, περιλαμβανομένων, μεταξύ άλλων, κατά περίπτωση:*

α) της ψευδωνυμοποίησης και της κρυπτογράφησης δεδομένων προσωπικού χαρακτήρα… (Ευρωπαϊκό Κοινοβούλιο, p. 66)

Σημαντικό είναι να αναφέρουμε πως στο σκεπτικό του ευρωπαϊκού κοινοβουλίου και του συμβουλίου της ευρωπαϊκής ένωσης για το σχεδιασμό του ΓΚΠΔ έχουν εκτιμηθεί πως τα δεδομένα προσωπικού χαρακτήρα που έχουν υποστεί ψευδωνυμοποίηση, η οποία θα μπορούσε να αποδοθεί σε φυσικό πρόσωπο με τη χρήση συμπληρωματικών πληροφοριών, θα πρέπει να θεωρούνται πληροφορίες σχετικά με ταυτοποιήσιμο φυσικό πρόσωπο. Επίσης η χρήση της ψευδωνυμοποίησης στα δεδομένα προσωπικού χαρακτήρα μπορεί να μειώσει τους κινδύνους για τα υποκείμενα των δεδομένων και να διευκολύνει τους υπευθύνους επεξεργασίας και τους εκτελούντες την επεξεργασία να τηρήσουν τις οικείες υποχρεώσεις περί προστασίας των δεδομένων. Το κίνητρο για τη λήψη μέτρων ψευδωνυμοποίησης είναι η επεξεργασία δεδομένων προσωπικού χαρακτήρα, με παράλληλη δυνατότητα μιας γενικής ανάλυσης, στο

πλαίσιο του ίδιου υπευθύνου επεξεργασίας, (Ευρωπαϊκό Κοινοβούλιο, p. 6). Υπάρχει αναλυτικό παράδειγμα στη σελίδα 22 της παρούσης εργασίας.



Εικόνα 1: Απεικόνιση ψευδωνυμοποίησης. Ιδίας επεξεργασίας

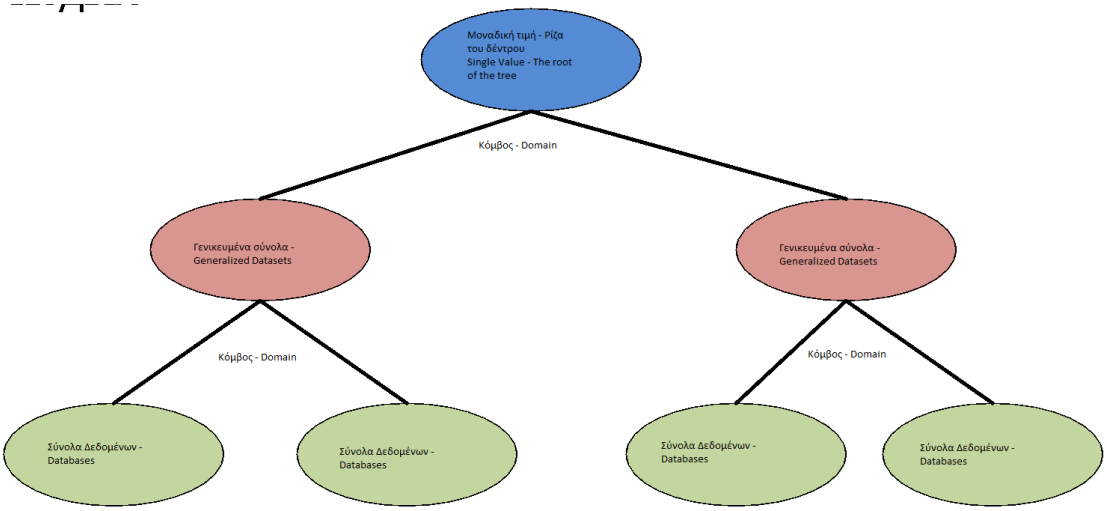
1.4. Ιεραρχίες γενίκευσης (Generalization hierarchy)

Η γενίκευση δεδομένων αναφέρεται στην διαδικασία απεικόνισης τιμών από ένα αρχικό πεδίο, έτσι ώστε διαφορετικές τιμές του αρχικού πεδίου να απεικονίζονται σε μία τιμή στο πεδίο προορισμού. Αυτό στην γενική περίπτωση επιτυγχάνεται με χρήση της ιεραρχίας γενίκευσης (generalization hierarchy) όπου κάθε τιμή του αρχικού πεδίου μπορεί να απεικονιστεί στο αμέσως γενικότερο επίπεδο και αυτή με την σειρά της στο αμέσως γενικότερο. Μπορούμε λοιπόν να σκεφτούμε την ιεραρχία γενίκευσης σαν ένα δέντρο που τα φύλλα απεικονίζονται στην τιμή του γονέα, αυτή στου δικού του γονέα πηγαίνοντας μέχρι την ρίζα του δέντρου, που σημασιολογικά αντιστοιχεί σε όλες τις τιμές. Γενίκευση μπορούμε να έχουμε τόσο σε ποσοτικές όσο και σε ποιοτικές τιμές. Οι ποσοτικές τιμές κατά την γενίκευσή τους αντιστοιχίζονται σε ένα διάστημα τιμών. Για παράδειγμα οι τιμές 32, 37 και 39 θα μπορούσαν να γενικευθούν στην τιμή “3*” που σημαίνει οποιαδήποτε τιμή από 30 ως 39 ή στην τιμή [31-40] κλπ. Η τιμή [31- 40] θα μπορούσε να γενικευθεί κι αυτή σε μια άλλη τιμή όπως “≤50”. Για τις ποιοτικές τιμές μπορούμε να θεωρήσουμε ως παράδειγμα μια ιεραρχία όπου οι Η.Π.Α και ο Καναδάς γενικεύονται στην τιμή “Βόρεια Αμερική”, η Βραζιλία και η Αργεντινή γενικεύονται στην τιμή “Νότια Αμερική” και με την σειρά τους οι δύο αυτές τιμές γενικεύονται στην τιμή “Αμερική”. Σε ότι αφορά την ανωνυμοποίηση, το πόσα επίπεδα θα

πρέπει να γενικοποιηθεί μια τιμή, εξαρτάται κυρίως από τον όγκο - συχνότητα εμφάνισης των διάφορων τιμών. Είναι σαφές ότι η ιεραρχία δημιουργείται ομαδοποιώντας τιμές με κάποιο ή κάποια κοινά χαρακτηριστικά. (Καρράς κα., 2014)

Οι ιεραρχίες γενίκευσης είναι ένα σύνολο κανόνων που καθορίζουν τον τρόπο με τον οποίο συγκεκριμένες τιμές πρέπει να αντικατασταθούν από πιο γενικές για την ανωνυμοποίηση των δεδομένων. Η βασική ιδέα εδώ είναι ότι οι αξίες που είναι αρκετά συγκεκριμένες ώστε να ταυτοποιηθούν (π.χ. ταχυδρομικός κώδικας κατοικίας) αντικαθίστανται από πιο γενικές (π.χ. ονόματα πόλεων) ώστε να μην μπορούν πλέον να αποκαλύψουν την ταυτότητα ενός ατόμου. Με άλλα λόγια, η συγκέντρωση των οιονεί αναγνωριστικών σε μεγαλύτερες ομάδες επιτυγχάνεται μέσω γενίκευσης, επομένως μειώνεται η ιδιαιτερότητα ενός χαρακτηριστικού αντικαθιστώντας μια συγκεκριμένη τιμή με μια πιο γενική. Για παράδειγμα, συγκεκριμένες ηλικιακές τιμές μπορούν να γενικευθούν σε ηλικιακές ομάδες (π.χ. η ομάδα 30-35 περιλαμβάνει τις τιμές ηλικίας 30, 31, 32, 33, 34, 35 ετών).(Marques & Bernardino, 2020)

Ένα παράδειγμα ιεραρχίας απεικονίζεται στα παρακάτω σχήμα :



Εικόνα 2: Απεικόνιση ιεραρχίας που καταλήγει σε μοναδική τιμή. Ιδίας επεξεργασίας.

Τα λογισμικά ανωνυμοποίησης (ενδεικτικά αναφέρουμε τα Amnesia , ARX κα) χρησιμοποιούν τις ιεραρχίες για να αντικαταστήσουν συγκεκριμένες τιμές με πιο γενικές έως ότου επιτευχθεί τέτοια γενικοποίηση ώστε να υπάρξει εγγύηση απορρήτου. Ένα χαρακτηριστικό των ιεραρχιών γενίκευσης είναι ότι όλοι οι κόμβοι οδηγούν σε έναν μόνο

κόμβο (ρίζα) (βλ. εν. 1.5). Αυτή η ιδιότητα εγγυάται στα λογισμικά ότι θα μπορούν να αντικαταστήσουν όλες τις τιμές με μια κοινή, εάν χρειαστεί.

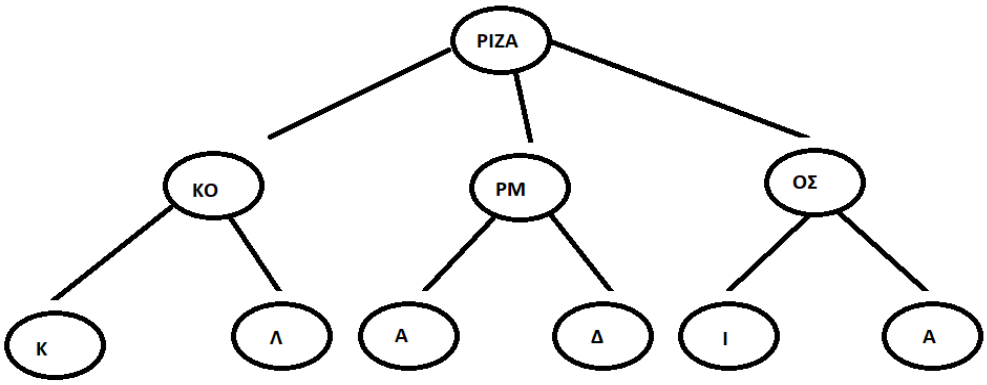
Οι ιεραρχίες γενίκευσης περιέχουν σημασιολογικές πληροφορίες που πρέπει να παρέχει ο χρήστης. Στην περίπτωση τομέων όπου οι σημασιολογικές πληροφορίες συνδέονται με μια συνολική σειρά (π.χ. αριθμοί, ημερομηνίες κ.λπ.), τα λογισμικά μπορούν να βοηθήσουν τον χρήστη να δημιουργήσει νέες ιεραρχίες. Οι ιεραρχίες μπορούν να αποθηκευτούν και να φορτωθούν ανάλογα το λογισμικό σε διάφορες μορφές (πχ .txt ή .csv). Επίσης ανάλογα το λογισμικό (όπως πχ το amnesia) προσφέρονται έτοιμες προς χρήση ιεραρχίες για ορισμένες σημαντικές οντολογίες του πραγματικού κόσμου όπως Διεθνής Στατιστική Ταξινόμηση Νόσων και Συναφών Προβλημάτων Υγείας (International Classification of Diseases - κώδικες ICD), ταχυδρομικοί κώδικες κλπ. (Amnesia software v. 1.3.2 (04/2023); International Classification of Diseases (ICD), accessed 2023).

1.5. Δέντρο και ρίζα του δέντρου (Tree and root)

Σύμφωνα με τον Buneman (1997) , ένας τρόπος αναπαράστασης των δεδομένων είναι η ενοποίησή τους σε κάποιου είδους δομή με επιτρεπόμενους κύκλους που μοιάζει με γραφικό ή δέντρο και θα αναφερόμαστε γενικά σε αυτά τα γραφήματα ως δέντρα.

Αυτά τα γραφήματα ονομάστηκαν δέντρα γιατί η ο τρόπος που παρουσιάζονται μοιάζει με δέντρο. Όπως τα δέντρα έχουν κλαδιά και ρίζες, έτσι και το γράφημα έχει κόμβους (nodes) με τον πρώτο από αυτούς να ονομάζεται ρίζα (root) του δέντρου και ακμές που είναι τα κλαδιά και ενώνουν κύκλο με κύκλο. (Συνδουκάς κα (προσπελάστηκε 2023)

Ως έννοιες δένδρων έχουμε δηλαδή, διασυνδεδεμένους κόμβους (nodes), με προγόνους και απογόνους, με κορυφή τη ρίζα (root) του δέντρου και στο τέλος τα φύλλα (leaves). (Στεφανιδάκης, 2016)



Εικόνα 3: Απεικόνιση Ρίζας του Δέντρου. Ιδίας επεξεργασίας.

1.6. k-Ανωνυμία / km–ανωνυμία (k-anonymization / km-anonymization)

Η k-ανωνυμία είναι μια πραγματική εγγύηση απορρήτου, αποδεδειγμένη και καθιερωμένη μαθηματικά. Εγγυάται ότι κάθε συνδυασμός τιμών οιονεί αναγνωριστικών μπορεί να αντιστοιχιστεί αδιάκριτα σε τουλάχιστον k άτομα. Για να επιτευχθεί αυτό, τα οιονεί αναγνωριστικά συγκεντρώνονται σε μια μεγαλύτερη ομάδα που περιέχει πληροφορίες που αντιστοιχούν σε οποιοδήποτε άτομο. Το k αναφέρεται στις εμφανίσεις κάθε συνδυασμού οιονεί αναγνωριστικών στο σύνολο δεδομένων. Εάν k=3, το σύνολο δεδομένων είναι 3-ανώνυμο, που σημαίνει ότι τουλάχιστον τρία άτομα μοιράζονται τις ίδιες τιμές οιονεί αναγνωριστικού.

Race	Birth	Gender	ZIP	Problem
Black	1965	m	0214*	short breath
Black	1965	m	0214*	chest pain
Black	1965	f	0213*	hypertension
Black	1965	f	0213*	hypertension
Black	1964	f	0213*	obesity
Black	1964	f	0213*	chest pain
White	1964	m	0213*	chest pain
White	1964	m	0213*	obesity
White	1964	m	0213*	short breath
White	1967	m	0213*	chest pain
White	1967	m	0213*	chest pain

Εικόνα 4: Παράδειγμα k-ανωνυμίας, L. Sweeney (2002).

Στην περίπτωση τιμών δεδομένων που είναι άσχετα με τον σκοπό της συλλογής τους, η k-ανωνυμία μπορεί να επιτευχθεί χρησιμοποιώντας την ολοκληρωτική απόκρυψη του χαρακτηριστικού (suppression). Ουσιαστικά αναφερόμαστε στην εξ ολοκλήρου αφαίρεση ενός χαρακτηριστικού που δεν θέλουμε να συμπεριλάβουμε ή είναι αδιάφορο για το σκοπό που μελετάται το σύνολο δεδομένων. (Amnesia software v. 1.3.2 (04/2023))

Σύμφωνα με την Δημοπούλου (2022), η μέθοδος της k-ανωνυμίας είναι πολύ συγκεκριμένη καθώς συνδυάζει και ομαδοποιεί προσωπικά δεδομένα με παρόμοια χαρακτηριστικά. Ένα σύνολο δεδομένων μπορεί να θεωρηθεί ως k-ανώνυμο όταν υπάρχει ένα σύνολο δεδομένων με προσωπικές πληροφορίες, με ίδιες τιμές χαρακτηριστικών που αντιστοιχούν σε τουλάχιστον k άτομα.

Η k-ανωνυμία παρέχει προστασία του απορρήτου διασφαλίζοντας ότι κάθε εγγραφή δεδομένων σχετίζεται με τουλάχιστον k=άτομα, ώστε να προστατεύονται τα δεδομένα ακόμη και αν συνδέονται ή συσχετίζονται με εξωτερικές διασυνδέσεις. (Sweeney, 2002)

Η ανωνυμία Km είναι μια πιο αδύναμη μορφή k-ανωνυμίας που είναι πιο κατάλληλη για δεδομένα υψηλών διαστάσεων. Όπως και στην k-ανωνυμία, ο αλγόριθμος λαμβάνει υπόψη έναν αριθμό n οιονεί αναγνωριστικών, αλλά τώρα περιορίζει την εγγύηση έναντι αντιπάλων που γνωρίζουν μόνο m από τα n οιονεί αναγνωριστικά (m << n). Με άλλα λόγια, ο αλγόριθμος ανωνυμοποίησης εγγυάται ότι κάθε συνδυασμός m οιονεί αναγνωριστικών εμφανίζεται τουλάχιστον k φορές στα σύνολα δεδομένων, ανεξάρτητα από τον συνολικό αριθμό n των οιονεί αναγνωριστικών. (Amnesia software v. 1.3.2 (04/2023))

Η διαφορά της km-ανωνυμίας είναι ότι διασφαλίζει το γεγονός ότι οποιοσδήποτε επιτιθέμενος γνωρίζει έως και m στοιχεία μιας εγγραφής ενός συνόλου δεδομένων, δεν μπορεί να χρησιμοποιήσει αυτή τη γνώση για να ταυτοποιήσει περισσότερα από k άτομα στο συγκεκριμένο σύνολο δεδομένων. (Gkountouna et al., 2014)

1.7. Δημογραφική στατιστική (Demographic statistic)

Η δημογραφία είναι η μελέτη με στατιστικές μεθόδους ανθρώπινων πληθυσμών, που περιλαμβάνει τη μέτρηση του μεγέθους, της ανάπτυξης και της μείωσής τους, των αναλογιών που ζουν σε μια περιοχή, γεννιούνται, παντρεύονται, έχουν παιδιά ή πεθαίνουν και τις σχετικές λειτουργίες της γονιμότητας, του γάμου και της θνησιμότητας. (Cox, 1976)

Η ανωνυμοποίηση των δημογραφικών στοιχείων μπορεί επίσης να πραγματοποιηθεί μέσω της γενίκευσης τους ενώ παράλληλα διατηρείται η χρηστικότητά τους στον τομέα της στατιστικής και επιστημονικής τους έρευνας. Σε αυτήν την περίπτωση, η k-ανωνυμία ενδέχεται να μην ισχύει για ολόκληρο το σύνολο δεδομένων. Ωστόσο, είναι εγγυημένη η κατανομή του πληθυσμού. Επομένως στην περίπτωση της δημογραφικής στατιστικής παραλλάσουμε το αρχικό σύνολο δεδομένων ειδικά όταν επιθυμούμε πληθυσμιακή ανωνυμοποίηση πχ 17/01/1982 -→παραλλαγή 3 ημερών -→15/01/1982 ή 19/01/1982 χωρίς να υποδεικνύουμε άτομα ή γεωγραφική ανωνυμοποίηση χωρίς να υποδεικνύουμε τοποθεσία. (Amnesia software v. 1.3.2 (04/2023))

ΑΡΧΙΚΟ ΣΥΝΟΛΟ			ΠΑΡΑΛΛΑΓΗ 3 ΗΜΕΡΩΝ ΠΡΙΝ			ΠΑΡΑΛΛΑΓΗ 3 ΗΜΕΡΩΝ ΜΕΤΑ	
ΑΤΟΜΟ	ΓΕΝΝΗΣΗ		ΑΤΟΜΟ	ΓΕΝΝΗΣΗ		ΑΤΟΜΟ	ΓΕΝΝΗΣΗ
ΒΑΙΟΣ	15/12		ΑΝΤΡΑΣ	13/12		ΑΝΤΡΑΣ	17/12
ΕΛΠΙΔΑ	16/12		ΓΥΝΑΙΚΑ			ΓΥΝΑΙΚΑ	
ΓΙΩΡΓΟΣ	17/12		ΑΝΤΡΑΣ			ΑΝΤΡΑΣ	
ΣΤΕΛΛΑ	15/12		ΓΥΝΑΙΚΑ	14/12		ΓΥΝΑΙΚΑ	18/12
ΤΑΣΟΣ	16/12		ΑΝΤΡΑΣ			ΑΝΤΡΑΣ	
ΓΙΩΤΑ	17/12		ΑΝΤΡΑΣ			ΑΝΤΡΑΣ	
ΔΗΜΗΤΡΗΣ	15/12		ΑΝΤΡΑΣ	15/12		ΑΝΤΡΑΣ	19/12
ΝΙΚΟΛΑΟΣ	16/12		ΓΥΝΑΙΚΑ			ΓΥΝΑΙΚΑ	
ΦΩΤΕΙΝΗ	17/12		ΓΥΝΑΙΚΑ			ΓΥΝΑΙΚΑ	

Εικόνα 5: Παραλλαγή τριών ημερών ημερομηνίας γέννησης. Ιδίας επεξεργασίας

Υπάρχουν χώρες που δημοσιεύουν τα μητρώα πληθυσμού που ενσωματώνουν την ταυτότητα των πολιτών μαζί με τα δημογραφικά τους στοιχεία. Αυτές οι πληροφορίες, οι οποίες συχνά διαδίδονται δημόσια, πωλούνται και μπορούν να χρησιμοποιηθούν ως σύνδεση με τα πραγματικά υποκείμενα.(Simi et al., 2017)

1.8. Κρυπτογράφηση (Encryption)

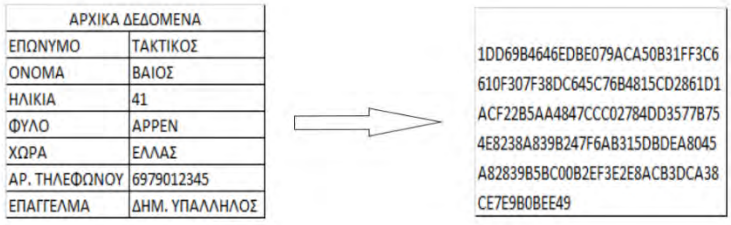
Κρυπτογραφία είναι η επιστήμη και η ικανότητα να γράφεις με μυστικότητα – κρατώντας τις πληροφορίες μυστικές. Κρυπτογράφηση είναι μία διεργασία με την οποία ένα μήνυμα (που ονομάζεται plaintext) μετατρέπεται σε ένα άλλο μήνυμα (που ονομάζεται ciphertext) χρησιμοποιώντας μια μαθηματική συνάρτηση (αλγόριθμος κρυπτογράφησης) και ένα ειδικό password κρυπτογράφησης, που ονομάζεται κλειδί. (Γκαδόλος, 1998)

Κρυπτογράφηση είναι η μετατροπή του συνόλου των δεδομένων σε ακατάληπτη – μη αναγνώσιμη μορφή. Η επεξεργασία των δεδομένων μπορεί να πραγματοποιηθεί μετά από την αποκρυπτογράφηση. (Σιουγλέ κα, (2019))

Σύμφωνα με το σκεπτικό του ΓΚΠΔ – GDPR: *Για τη διατήρηση της ασφάλειας και την αποφυγή της επεξεργασίας κατά παράβαση του παρόντος κανονισμού, ο υπεύθυνος επεξεργασίας ή ο εκτελών την επεξεργασία θα πρέπει να αξιολογεί τους κινδύνους που ενέχει η επεξεργασία και να εφαρμόζει μέτρα για τον μετριασμό των εν λόγω κινδύνων, όπως για παράδειγμα μέσω κρυπτογράφησης…*(Ευρωπαϊκό Κοινοβούλιο, n.d., p. 21)

Επίσης, όπως και στην περίπτωση της ψευδωνυμοποίησης για την ασφάλεια των προσωπικών δεδομένων αναφέρεται η ύπαρξη κατάλληλων εγγυήσεων, που μπορεί να περιλαμβάνουν κρυπτογράφηση ή ψευδωνυμοποίηση, όπως επίσης ο υπεύθυνος επεξεργασίας και ο εκτελών την επεξεργασία να εφαρμόζουν κατάλληλα τεχνικά και οργανωτικά μέτρα προκειμένου να διασφαλίζεται το κατάλληλο επίπεδο ασφαλείας έναντι των κινδύνων, περιλαμβανομένων μεταξύ άλλων κατά περίπτωση: α) την ψευδωνυμοποίησης και τη κρυπτογράφηση δεδομένων προσωπικού χαρακτήρα, …. (Ευρωπαϊκό Κοινοβούλιο, n.d., p. 47,66)

Η κρυπτογράφηση καθιστά ολόκληρα τα δεδομένα «μη αναγνώσιμα» και ουσιαστικά δεν μπορεί να γίνει κάποια στατιστική ανάλυση επί των κρυπτογραφημένων δεδομένων ή να αποκαλυφθεί κάποια άλλη πληροφορία γιατί πρέπει υποχρεωτικά να αποκρυπτογραφηθούν πρώτα με το αντίστοιχο κλειδί αποκρυπτογράφησης. Αν και υπάρχουν κρυπτογραφικές τεχνικές που «επιτρέπουν» κάποιες πράξεις επί κρυπτογραφημένων μηνυμάτων (πχ. Ομομορφική κρυπτογραφία), εν τούτοις η διαφορά ψευδωνυμοποιημένων με κρυπτογραφημένων δεδομένων είναι καταφανής. (Λιμνιώτης, 2018)



ΑΡΧΙΚΑ ΔΕΔΟΜΕΝΑ	
ΕΠΩΝΥΜΟ	ΤΑΚΤΙΚΟΣ
ΟΝΟΜΑ	ΒΑΙΟΣ
ΗΛΙΚΙΑ	41
ΦΥΛΟ	ΑΡΡΕΝ
ΧΩΡΑ	ΕΛΛΑΣ
ΑΡ. ΤΗΛΕΦΩΝΟΥ	6979012345
ΕΠΑΓΓΕΛΜΑ	ΔΗΜ. ΥΠΑΛΛΗΛΟΣ

1DD69B4646EDBE079ACA50B31FF3C6
610F307F38DC645C76B4815CD2861D1
ACF22B5AA4847CCC02784DD3577B75
4E8238A839B247F6AB315DBDEA8045
A82839B5BC00B2EF3E2E8ACB3DCA38
CE7E9B0BEE49

Εικόνα 6: Απεικόνιση κρυπτογράφησης. Ιδίας επεξεργασίας

1.9. Συγκάλυψη (Masking)

Η κάλυψη αναφέρεται σε μια μέθοδο αλλαγής χαρακτήρων σε επιλεγμένο χαρακτηριστικό σε διαφορετικό χαρακτήρα, καθιστώντας τη μεταβλητή ασύλληπτη. Οποιοσδήποτε αριθμητικός αριθμός από το 1-9 θα αντικατασταθεί με το 1 και κάθε πεζό a-z θα αντικατασταθεί με το z και κάθε κεφαλαίο A-Z θα αντικατασταθεί με το Z. Ο πρώτος χαρακτήρας, ο αριθμός 0 και ο ειδικός χαρακτήρας θα διατηρηθούν ως η αρχική τιμή. Το πρόβλημα της κάλυψης είναι ότι θα καταναλώσει περισσότερους πόρους για τον έλεγχο και την αλλαγή της τιμής, αλλά τελικά τα δεδομένα είναι άχρηστα για έρευνα. Στην κατάσταση κάλυψης μπορούμε να χρησιμοποιήσουμε καταστολή που δεν θα ελέγξει την τιμή αλλά θα αλλάξει όλη την τιμή σε ορισμένα σύμβολα ενώ θα αρχειοθετήσει το ίδιο αποτέλεσμα με την κάλυψη με μεγαλύτερη αποτελεσματικότητα. (Murthy et al., 2019)

Η συγκάλυψη είναι μια μέθοδος ψευδο-ανωνυμοποίησης που αναφέρεται στην απόκρυψη ορισμένων πληροφοριών στο σύνολο δεδομένων χρησιμοποιώντας εναλλακτικούς χαρακτήρες. Οι τεχνικές συγκάληψης χρησιμοποιούνται ευρέως για την απόκρυψη τμημάτων των αριθμών πιστωτικών καρτών κατά τη διάρκεια διαδικασιών και πληρωμών πιστωτικών καρτών.

Ωστόσο, τα εναπομείναντα δεδομένα ταυτοποίησης (ημερομηνία γέννησης, ταχυδρομικός κώδικας, φύλο, οικογενειακή κατάσταση, κ.λπ.) θα μπορούσαν να συνδυαστούν για τον επαναπροσδιορισμό των προσώπων και τη διακύβευση του απορρήτου τους. Έχει αποδειχτεί ότι το τριπλό (ημερομηνία γέννησης, φύλο, ταχυδρομικός κώδικας) είναι αρκετό για να

προσδιορίσει μοναδικά τουλάχιστον το 87% των πολιτών των ΗΠΑ σε δημόσια διαθέσιμα σύνολα δεδομένων. Τέτοια δεδομένα αναγνώρισης ονομάζονται επίσης οιονεί αναγνωριστικά ή έμμεσα αναγνωριστικά. (Amnesia software v. 1.3.2 (04/2023)); Sweeney, 2000; United Nations Development Group, 2017).

Το σκεπτικό της ψευδωνυμοποίησης είναι ότι δεν είναι πάντα δυνατό να καταστούν τα δεδομένα ανώνυμα και ταυτόχρονα να διατηρηθούν οι απαραίτητες πληροφορίες για τους σκοπούς της επεξεργασίας, ειδικά για επιστημονικούς, στατιστικούς ή ιστορικούς σκοπούς. "Αντικατάσταση ενός χαρακτηριστικού με ένα άλλο" σημαίνει ότι η ταυτότητα του υποκειμένου καλύπτεται (το υποκείμενο δεν προσδιορίζεται άμεσα) αλλά ταυτόχρονα παραμένει η ικανότητα ταυτοποίησής του (ένα έμμεσα αναγνωρίσιμο υποκείμενο). Αυτή η συνθήκη είναι χρήσιμη όταν:

– Τα δεδομένα που σχετίζονται με το ίδιο θέμα πρέπει να συλλέγονται αλλά χωρίς να είναι γνωστή η ταυτότητά του.

– Υπάρχει ανάγκη να διατηρηθεί η δυνατότητα επαναπροσδιορισμού του υποκειμένου για να διασφαλιστεί ότι όλες οι πληροφορίες που διατηρούνται αποδίδονται σε αυτόν/αυτήν. (Bolognini & Bistolfi, 2017)

Αντίστοιχα στην περίπτωση γενικών Δεδομένων ή Δεδομένων Γενικού Σκοπού, δεν έχουμε γνώση για το είδος της ανάλυσης που πρέπει να πραγματοποιηθεί από τον αποδέκτη των δεδομένων για την έρευνα. Αυτή είναι η συνήθης περίπτωση κατά την οποία τα δεδομένα δημοσιεύονται μέσω διακομιστή για μελλοντική χρήση, περιλαμβάνει επίσης την περίπτωση που τα δεδομένα μεταφέρονται σε έναν επεξεργαστή δεδομένων ή σε έναν ερευνητή για την ανάλυσή τους, καθώς συνήθως δεν γνωρίζουμε ποιος αλγόριθμος θα εφαρμοστεί στα δεδομένα. Για το σκοπό αυτό, έχουν αναπτυχθεί μέθοδοι ανωνυμοποίησης, γνωστές και ως μέθοδοι κάλυψης (Torra & Navarro-Arribas, 2016).

ΑΡΧΙΚΑ ΔΕΔΟΜΕΝΑ	
ΕΠΩΝΥΜΟ	ΤΑΚΤΙΚΟΣ
ΟΝΟΜΑ	ΒΑΙΟΣ
ΗΛΙΚΙΑ	41
ΦΥΛΟ	ΑΡΡΕΝ
ΧΩΡΑ	ΕΛΛΑΣ
Τ.Κ.	38333
ΑΡ. ΤΗΛΕΦΩΝΟΥ	6979012345
e-mail	vaitak@yahoo.gr

ΕΦΑΡΜΟΓΗ ΣΥΓΚΑΛΥΨΗΣ	
ΕΠΩΝΥΜΟ	ΤΑΚΤΙΚΟΣ
ΟΝΟΜΑ	ΒΑΙΟΣ
ΗΛΙΚΙΑ	**
ΦΥΛΟ	ΑΡΡΕΝ
ΧΩΡΑ	ΕΛΛΑΣ
Τ.Κ.	****3
ΑΡ. ΤΗΛΕΦΩΝΟΥ	*********5
e-mail	******@yahoo.gr

Εικόνα 7:Απεικόνιση τεχνικής συγκάλυψης. Ιδίας επεξεργασίας

1.10. Διαφορικό απόρρητο (Differential privacy)

Το διαφορικό απόρρητο είναι μία αυστηρή μαθηματική εφαρμογή. Στην απλούστερη μορφή της, πρόκειται για έναν αλγόριθμο που αναλύει ένα σύνολο δεδομένων και υπολογίζει στατιστικά δεδομένα σχετικά με αυτό (όπως πχ ο μέσος όρος των δεδομένων, η διακύμανση, η διάμεσος, ο τρόπος λειτουργίας κ.λπ.). Ένας τέτοιος αλγόριθμος λέγεται ότι είναι διαφορικά ιδιωτικός, εάν παρατηρώντας τα τελικά αποτελέσματα από την εφαρμογή του, δεν μπορούμε να τα επαληθεύσουμε με τα δεδομένα που χρησιμοποιήθηκαν στο αρχικό σύνολο δεδομένων. Η εγγύηση ανωνυμοποίησης που προσφέρει η εφαρμογή ενός διαφορικά ιδιωτικού αλγορίθμου είναι ότι δεν μπορεί να διακριθεί το δεδομένο εισαγωγής από το γενικό τελικό σύνολο δεδομένων. Ο αλγόριθμος εφαρμόζει θόρυβο στα αρχικά δεδομένα και μπορεί να εξάγει μια βάση δεδομένων που περιέχει πληροφορίες χωρίς να μπορεί να επιβεβαιωθεί η αρχική βάση δεδομένων. Αυτή η εγγύηση ισχύει για οποιοδήποτε άτομο και οποιοδήποτε σύνολο δεδομένων. (University of North Alambama, accessed 04/2023).

Το διαφορικό απόρρητο είναι ένας μαθηματικά αυστηρός ορισμός του απορρήτου προσαρμοσμένος στην ανάλυση μεγάλων συνόλων δεδομένων και εξοπλισμένος με ένα επίσημο μέτρο απώλειας απορρήτου. Επιπλέον, οι διαφορικά ιδιωτικοί αλγόριθμοι λαμβάνουν ως είσοδο μια παράμετρο, που συνήθως ονομάζεται ε, η οποία περιορίζει την επιτρεπόμενη απώλεια απορρήτου σε οποιαδήποτε εκτέλεση του αλγορίθμου και προσφέρει μια συγκεκριμένη αντιστάθμιση ιδιωτικότητας και χρησιμότητας. Ένα από τα δυνατά σημεία του διαφορικού απορρήτου είναι η ικανότητα να συλλογίζεται η σωρευτική απώλεια απορρήτου σε

πολλαπλές αναλύσεις, δεδομένων των τιμών του ε που χρησιμοποιούνται σε κάθε μεμονωμένη ανάλυση. (Dwork & Rothblum, 2016)

Η Apple υιοθέτησε και ανέπτυξε περαιτέρω μια τεχνική γνωστή στον ακαδημαϊκό κόσμο ως τοπικό διαφορικό απόρρητο με σκοπό να αποκτήσει μια εικόνα για το τι κάνουν πολλοί χρήστες της Apple, συμβάλλοντας παράλληλα στη διατήρηση του απορρήτου των μεμονωμένων χρηστών. Είναι μια τεχνική που επιτρέπει στην Apple να μαθαίνει για την κοινότητα των χρηστών χωρίς να μαθαίνει για άτομα στην κοινότητα. Το διαφορικό απόρρητο μετατρέπει τις πληροφορίες που μοιράζονται με την Apple προτού φύγουν από τη συσκευή του χρήστη, έτσι ώστε η Apple να μην μπορεί ποτέ να αναπαράγει τα αληθινά δεδομένα. (*Differential Privacy A Privacy-Preserving System*, (accessed 03/2023)

Η Google χρησιμοποιεί το διαφορικό απόρρητο ως μια προηγμένη τεχνολογία ανωνυμοποίησης που επιτρέπει την απόκτηση πληροφοριών από δεδομένα χωρίς να διακυβεύεται η ανωνυμία των χρηστών. Η έρευνα διήρκησε πάνω από μια δεκαετία δημιουργώντας τη μεγαλύτερη βιβλιοθήκη διαφορικών αλγορίθμων απορρήτου στον κόσμο και δημιουργήθηκε μια βιβλιοθήκη ανοιχτού κώδικα για να βοηθηθούν οι οργανισμοί και να εφαρμόσουν εύκολα τις ίδιες προστασίες απορρήτου στα δεδομένα τους. (Differential Privacy Accounting by Connecting the Dots – Google AI Research, (accessed 03/2023)

1.11. Ομοσπονδιακή μάθηση (Federated learning)

Εφαρμογές και τεχνολογίες για κινητές συσκευές όπως το Διαδίκτυο των πραγμάτων (Internet of Things (IoT)) και τα κοινωνικά δίκτυα έχουν οδηγήσει σε εκπληκτική ανάπτυξη της κίνησης δεδομένων και των φορητών συσκευών, γεγονός που φέρνει προκλήσεις για την επικοινωνία δεδομένων και την ασφαλή χρήση υπολογιστών στην τεχνολογία 6G. Οι παραδοσιακές μέθοδοι διατήρησης της ιδιωτικής ζωής βασίζονται κυρίως σε παραμόρφωση δεδομένων και κρυπτογράφηση δεδομένων, αλλά αυτές οι μέθοδοι χρησιμοποιούνται με δυσκολία λόγω των ανεπαρκών επιπτώσεων προστασίας και της πολύπλοκης εφαρμογής. Για να βελτιώσει την προστασία της ιδιωτικής ζωής, η Dwork πρότεινε την έννοια του διαφορικού απορρήτου το 2006. Το απόρρητο μπορεί να αντισταθεί σε διάφορες νέες επιθέσεις με την παραδοχή της μέγιστης γνώσης του υποβάθρου. Το 2016, η Google πρότεινε την ομοσπονδιακή μάθηση ως ένα ασφαλές πλαίσιο μηχανικής εκμάθησης για την επίλυση του προβλήματος

απορρήτου δεδομένων σε υπολογιστές κινητών συσκευών μεγάλης κλίμακας. Η ομοσπονδιακή μάθηση είναι ένα από τα πιο δημοφιλή πλαίσια Secure Multiparty Computing (SMC) στη μηχανική εκμάθηση, το οποίο επιτρέπει σε πολλούς πελάτες να εκπαιδεύουν από κοινού μοντέλα σε έναν κεντρικό διακομιστή και έχει εφαρμοστεί σε σενάρια επικοινωνίας 5G. Αυτό παρέχει επίσης μια ασφαλή και αποτελεσματική μέθοδο υπολογισμού για επικοινωνία 6G (Wu et al., 2022)

Η ομοσπονδιακή μάθηση έχει τη δυνατότητα να ενεργοποιεί λειτουργίες πρόβλεψης σε smartphone κατά την επικοινωνία της με τον κεντρικό διακομιστή, χωρίς να μειώνει την εμπειρία του χρήστη ή να διαρρέει προσωπικές πληροφορίες. (Li et al., 2020)

Η ομοσπονδιακή μάθηση είναι ένα παράδειγμα μάθησης που επιδιώκει να αντιμετωπίσει το πρόβλημα της διακυβέρνησης και της ιδιωτικής ζωής των δεδομένων μέσω της εκπαίδευσης αλγορίθμων συνεργατικά χωρίς ανταλλαγή των ίδιων των δεδομένων.(Rieke et al., 2020)

Η ομοσπονδιακή μάθηση είναι ένα παράδειγμα μηχανικής μάθησης που προτείνεται ως πιθανή απάντηση στην απαίτηση διατήρησης του απορρήτου των δεδομένων, μαζί με μια κατανεμημένη προσέγγιση για την αντιμετώπιση της τοπικής και παγκόσμιας μάθησης. Επίσης στοχεύει στη δημιουργία ενός συλλογικά εκπαιδευμένου παγκόσμιου μοντέλου μάθησης χωρίς να μοιράζεται τα δεδομένα που ανήκουν στις κατανεμημένες πηγές δεδομένων. Η μηχανική μάθηση είναι ευάλωτη σε επιθέσεις από κακόβουλο χρήστη που δημιουργούν δυσλειτουργία των εφαρμογών που στηρίζονται σε μοντέλα μηχανικής μάθησης. Ομοίως, η ομοσπονδιακή μάθηση εκτίθεται στον ίδιο κίνδυνο, καθώς είναι μια συγκεκριμένη ρύθμιση μηχανικής εκμάθησης. Συνεπώς αδύναμο σημείο της ομοσπονδιακής μάθησης είναι η έκθεση σε επιθέσεις αντιπάλου που μπορεί να παραβιάζουν την ακεραιότητα του μοντέλου εκμάθησης ή το απόρρητο των δεδομένων.(Rodríguez-Barroso et al., 2023; Κουδέρη, 2020)

1.12 Σύνολα μικροδεδομένων (Microdata datasets)

Σύμφωνα με τη Eurostat τα μικροδεδομένα είναι σύνολα αρχείων που περιέχουν πληροφορίες για άτομα, νοικοκυριά ή επιχειρήσεις. Χρησιμοποιούνται σε επίσημα στατιστικά στοιχεία για την παραγωγή συγκεντρωτικών πληροφοριών, συνήθως σε μορφή πίνακα.

Η πρόσβαση σε εμπιστευτικά μικροδεδομένα περιορίζεται για την προστασία της ανωνυμίας ιδιωτών ή επιχειρήσεων. Η πρόσβαση στα μικροδεδομένα παρέχεται μόνο για επιστημονικούς σκοπούς.

Για να μπορέσει κάποιος οργανισμός να έχει πρόσβαση σε αυτά τα δεδομένα πρέπει πρώτα να αιτηθεί την αναγνώρισή ως ερευνητικό φορέα (όπως πχ πανεπιστήμια, ερευνητικά ιδρύματα ή ερευνητικά τμήματα σε μια δημόσια διοίκηση, τράπεζες, στατιστικά ιδρύματα κ.λπ.), με χρονική διάρκεια της διαδικασίας τις τέσσερις εβδομάδες, ενώ η διαδικασία αίτησης πρόσβασης σε σύνολα μικροδεδομένων διαρκεί οκτώ έως δέκα εβδομάδες.

Στο τέλος της περιόδου πρόσβασης πρέπει: 1) να καταστρέφουν τυχόν πρωτότυπα αρχεία επιστημονικής χρήσης και 2) να σταλούν στην Eurostat τα αποτελέσματα της έρευνάς που πραγματοποιήθηκε. (*Overview - Microdata - Eurostat*, (accessed 05/2023)

ΚΕΦΑΛΑΙΟ 2

Πραγματικά συμβάντα διαρροής δεδομένων

2.1.Το παράδειγμα NETFLIX

Η επιτυχία της εταιρίας Netflix βασίζεται στη χρήση του one-to-one marketing με τη μορφή της «εξατομίκευσης» και μιας μεγάλης βάσης δεδομένων που στηρίζεται όχι μόνο στις αξιολογήσεις των πελατών της για τα τηλεοπτικά της προγράμματα, αλλά συλλέγει δεδομένα από την αλληλεπίδραση και ανταπόκριση των πελατών της σε μια τηλεοπτική εκπομπή. (Γιανναντίδης, 2022)

Στις 2 Οκτωβρίου του 2006, η εταιρία κυκλοφόρησε ένα μεγάλο σύνολο δεδομένων αξιολογήσεων ταινιών και προκάλεσε τις κοινότητες εξόρυξης δεδομένων, μηχανικής μάθησης και επιστήμης υπολογιστών, να αναπτύξουν συστήματα που θα μπορούσαν να υπερβούν την ακρίβεια της σύστασης επιλογών θεαμάτων στους πελάτες της. (Σύστημα (Cinematch) κατά 10%). Το Netflix παρείχε 100.480.507 βαθμολογίες (σε κλίμακα από 1 έως 5 αστέρια) μαζί με τις ημερομηνίες τους από 480.189 **τυχαία επιλεγμένους, ανώνυμους συνδρομητές σε** 17.770

τίτλους ταινιών, καθιστώντας το μακράν τις μεγαλύτερες βαθμολογίες που διατίθεται στην ερευνητική κοινότητα δεδομένων συστημάτων συστάσεων επιλογών. Προκειμένου να καταστήσει την πρόκληση πιο ενδιαφέρουσα, η εταιρεία ανακοίνωσε ότι θα απονείμει ένα Μεγάλο Βραβείο 1 εκατομμυρίου δολαρίων στην πρώτη ομάδα που θα μπορούσε να επιτύχει αυτόν τον στόχο. (Amatriain & Basilico, 2015)

Το βραβείο Netflix έθεσε τα φώτα της δημοσιότητας στην περιοχή των συστημάτων προτάσεων και στην αξία της δημιουργίας εξατομικευμένων προτάσεων από δεδομένα χρηστών. Το έκανε παρέχοντας έναν σαφή ορισμό του προβλήματος που επέτρεψε σε χιλιάδες ομάδες να επικεντρωθούν στη βελτίωση μιας μεμονωμένης μέτρησης. Ενώ αυτό ήταν μια απλοποίηση του προβλήματος των συστάσεων, πολλά πολύτιμα διδάγματα αντλήθηκαν. (Ampazis, 2010)

Ερωτηθείσα η Netflix αν στα ανωτέρω δημοσιευμένα δεδομένα υπάρχουν στοιχεία πελατών η εταιρία απάντησε : *"Όχι, όλες οι πληροφορίες αναγνώρισης πελατών έχουν αφαιρεθεί. το μόνο που μένει είναι βαθμολογίες και ημερομηνίες. Αυτό ακολουθεί την πολιτική απορρήτου μας, την οποία μπορείτε να διαβάσετε εδώ. Ακόμα κι αν, για παράδειγμα, γνωρίζατε όλες τις δικές σας αξιολογήσεις και τις ημερομηνίες τους, πιθανότατα δεν θα μπορούσατε να τις προσδιορίσετε αξιόπιστα στα δεδομένα, επειδή συμπεριλήφθηκε μόνο ένα μικρό δείγμα (λιγότερο από το ένα δέκατο του πλήρους συνόλου δεδομένων μας) και αυτά τα δεδομένα υπόκεινται σε όχληση. Φυσικά, δεδομένου ότι γνωρίζετε όλες τις δικές σας αξιολογήσεις, είναι αυτό πραγματικά πρόβλημα απορρήτου;"*

Το αποτέλεσμα ήταν εφαρμόζοντας διάφορους αλγόριθμους και με διασταυρούμενη συσχέτιση με την βάση δεδομένων ταινιών IMDB να ταυτοποιηθούν ανώνυμες αξιολογήσεις ταινιών της Netflix με τους ίδιους τους πελάτες. (Narayanan & Shmatikov, 2008)

2.2 Η περίπτωση του GRAVATAR

Το Gravatar είναι μια ευρέως χρησιμοποιούμενη υπηρεσία για την παροχή μοναδικών προσωπικών ειδώλων εικόνας παγκοσμίως. Το είδωλο ενός χρήστη είναι μια μοναδική εικόνα συνδεδεμένη με την ηλεκτρονική διεύθυνση (email) του χρήστη. Χρησιμοποιείται τεχνολογία κρυπτογράφησης (MD5 Hash της Java) (Wolf et al., 2017) για τη μοναδικότητα της σύνδεσης εικόνας με την ηλεκτρονική διεύθυνση του χρήστη. Η διαρροή ηλεκτρονικών διευθύνσεων έχει ως ακολούθως:

- Το 2008 ανακτήθηκαν περίπου 10% των ηλεκτρονικών διευθύνσεων της υπηρεσίας
- Το 2013 70% των ηλεκτρονικών διευθύνσεων
- Το 2019 παρουσιάστηκε βήμα βήμα η διαδικασία ανάκτησης
- Το 2020 παρουσιάστηκαν περίπου 20% πραγματικών ηλεκτρονικών διευθύνσεων

(Rodwald, 2021)

2.3 Η Περίπτωση του AOL no. 4417749

Η America Online (AOL) ήταν μία από τις μεγαλύτερες συνδρομητικές υπηρεσίες παροχής διαδικτύου στις ΗΠΑ. Ήταν από τις πρώτες εταιρίες που δημιούργησε μηνύματα μέσω διαδικτύου ακόμη από το 1989. (*AOL | American Company | Britannica*, n.d.). Στο αποκορύφωμά της η AOL είχε κεφαλαιοποίηση αγοράς άνω των 200 δισεκατομμυρίων δολαρίων (*How AOL Dominated the Internet of the '90s and Let It Slip Away*, n.d.). Τον Ιανουάριο του 2016 μαθεύτηκε πως η AOL, η Microsoft και η Yahoo ο έστειλε ερωτήματα αναζήτησης στο Διαδίκτυο μιας εβδομάδας από εκατομμύρια Αμερικανούς σε δικηγόρους του Υπουργείου Δικαιοσύνης. Οι δικηγόροι εξέδωσαν κλητεύσεις για τα δεδομένα που θα χρησιμοποιηθούν για τη δοκιμή λογισμικού φιλτραρίσματος σε σχέση με τον αγώνα του Υπουργείου να περιορίσει την έκθεση των παιδιών σε διαδικτυακή πορνογραφία.(Hillyard & Gauen, 2007)

Αποτέλεσμα αυτής της διαδικασίας ήταν ένας χρήστης με τον αριθμό 4417749 και χωρίς ιδιαίτερη έρευνα αφού ακολουθήθηκαν τα ίχνη των αναζητήσεων να οδηγηθούν στα πλήρη στοιχεία και εντοπισμό του ίδιου του χρήστη. (Barbaro et al., 2006)

2.4 Η περίπτωση του ηλεκτρονικού εισιτηρίου του ΟΑΣΑ

Ένα βασικό ερώτημα που θέτει το ηλεκτρονικό εισιτήριο είναι αν κινδυνεύει η ιδιωτικότητα των επιβατών που το χρησιμοποιούν (Μαρτσούκου κα, 2018). Το 2017 μία από τις καινοτομίες του Οργανισμού Αστικών Συγκοινωνών Αθηνών (ΟΑΣΑ) ήταν η ανακοίνωση της έναρξης ισχύος του ηλεκτρονικού εισιτηρίου, την διαδικασία του οποίου σταμάτησε η Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα.

Το πρόβλημα στην συγκεκριμένη περίπτωση ήταν ότι το ηλεκτρονικό εισιτήριο απαιτεί ταυτοποίηση για την έκδοση των εισιτηρίων των κατόχων – επιβατών, γεγονός που επιστρέφει αποτελέσματα σχετικά με:

- τα δρομολόγια,
- την οικονομική κατάσταση των επιβατών
- το ΑΜΚΑ
- Ημ/νια γέννησης
- Αριθμός τηλεφώνου
- Φωτογραφία σε ψηφιακή μορφή

(Κατσαβριά, 2018)

Αποτέλεσμα ήταν οι γνωμοδοτήσεις της ΑΠΔΠΧ με αριθμούς 01/2017 και 04/2017 με τις οποίες τα προσωπικά δεδομένα για το ηλεκτρονικό εισιτήριο επεξεργάζονται μόνο υπό τη μορφή προσωποποιημένης κάρτας, ενώ η επεξεργασία του δεν αφορά τις μη προσωποποιημένες κάρτες (ΑΠΔΠΧ Γνωμοδότηση 1/2017; ΑΠΔΠΧ Γνωμοδότηση 4/2017)

2.5 Πρακτικό παράδειγμα

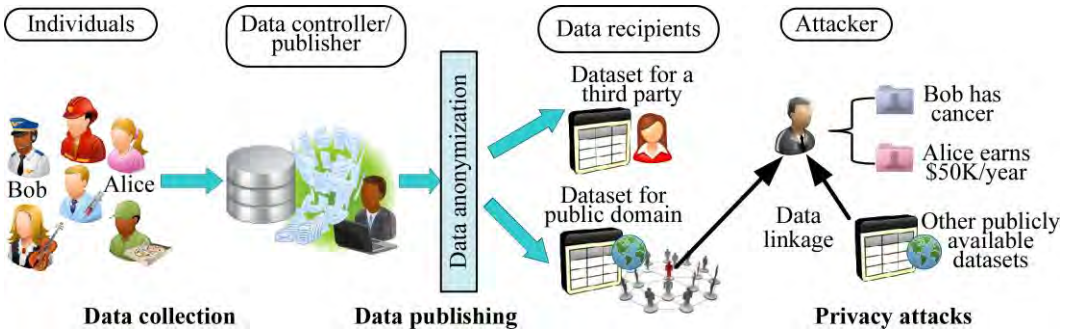
Πριν προχωρήσουμε καλό θα ήταν να αναφέρουμε ένα πρακτικό παράδειγμα διαρροής απόρρητων δεδομένων. Λαμβάνουμε υπόψη το πρόβλημα της δημοσίευσης δεδομένων με καθορισμένες - συγκεκριμένες τιμές, διατηρώντας παράλληλα το απόρρητο των ατόμων που σχετίζονται με αυτά.

Σκεφτόμαστε μια βάση δεδομένων Α, η οποία αποθηκεύει σε σχηματισμό αντικείμενα που αγοράστηκαν σε ένα φαρμακείο από διάφορους πελάτες. Παρατηρούμε ότι η απευθείας δημοσίευση του Α μπορεί να οδηγήσει στην αποκάλυψη της ταυτότητας του ατόμου που σχετίζεται με μια συγκεκριμένη συναλλαγή, εάν ο υποκλοπέας έχει μερική γνώση των αντικειμένων που αγόρασε αυτό το άτομο. Για παράδειγμα, ας υποθέσουμε ότι ο Βάιος πήγε στο φαρμακείο μια συγκεκριμένη ημέρα και αγόρασε ένα σετ ειδών όπως βιταμίνες, φάρμακα για πονοκέφαλο, φάρμακα για covid19. Ας υποθέσουμε επίσης ότι ορισμένα από τα αντικείμενα που αγόρασε ο Βάιος ήταν εμφανή, πάνω από τα ψώνια του (π.χ. φάρμακα για ίωση) και παρατηρήθηκαν από τον γείτονά του Χρήστο, ενώ και οι δύο βρίσκονταν στο ίδιο λεωφορείο. Ο Βάιος δεν θα ήθελε ο Χρήστος να μάθει τα υπόλοιπα αντικείμενα που αγόρασε.

Ωστόσο, εάν το φαρμακείο αποφασίσει να δημοσιεύσει τις συναλλαγές του και υπάρχει μόνο μία συναλλαγή που περιέχει βιταμίνες και φάρμακα για covid19, ο Χρήστος μπορεί αμέσως να συμπεράνει ότι αυτή η συναλλαγή αντιστοιχεί στον Βάιο και μπορεί να μάθει το πλήρες περιεχόμενο της τσάντας αγορών του. Αυτό το παράδειγμα τονίζει την ανάγκη μετατροπής της αρχικής βάσης δεδομένων συναλλαγών Α σε βάση δεδομένων Α' πριν από τη δημοσίευση, προκειμένου να αποφευχθεί η συσχέτιση συγκεκριμένων συναλλαγών με ένα συγκεκριμένο άτομο ή γεγονός.

Στην πράξη, αναμένουμε από τον υποκλοπέα να έχει μόνο μερική γνώση για τις συναλλαγές (διαφορετικά, θα υπήρχαν ελάχιστες ευαίσθητες πληροφορίες για απόκρυψη). Από την άλλη πλευρά, δεδομένου ότι η γνώση για τα δεδομένα από τον υποκλοπέα δεν είναι γνωστή στον εκδότη των δεδομένων, είναι λογικό να οριστεί ένα γενικό μοντέλο για το απόρρητο, το οποίο προστατεύει από τους αντιπάλους που έχουν γνώση περιορισμένη σε ένα επίπεδο, που εκφράζεται ως παράμετρος του μοντέλου. (Ayala-Rivera et al., 2014; Terrovitis et al., 2011)

Η ανωνυμοποίηση στην πράξη παρουσιάζεται πολύ επεξηγηματικά στην παρακάτω εικόνα:



Εικόνα 8: Ayala-Rivera et al.(2014). Η ανωνυμοποίηση στην πράξη.

ΚΕΦΑΛΑΙΟ 3

Επισκόπηση του λογισμικού ανωνυμοποίησης δεδομένων Amnesia

Το Amnesia πρόκειται για ένα ελληνικό, δωρεάν λογισμικό, ανοικτού κώδικα ανωνυμοποίησης δεδομένων, που αναπτύχθηκε από το Πανεπιστήμιο Θεσσαλίας και δημοσιεύτηκε για πρώτη φορά το 2019. Το λογισμικό παρέχει στο χρήστη μια σχεδόν αυτοματοποιημένη διαδικασία ανωνυμοποίησης συνόλου δεδομένων, με πλήρη καθοδήγηση αλλά με δυνατότητα παρέμβασης του χρήστη σε κάθε επίπεδο, για δεδομένα σε πίνακα και καθορισμένες τιμές (καθώς και συνδυασμούς αυτών). Μέρη της βάσης αλγορίθμων του βασίζονται στο λογισμικό ARX (βλ. εν. 4.1). (Haber et al., 2022)

Το Amnesia είναι ένα λογισμικό που επιτρέπει στους χρήστες να ανωνυμοποιούν τα δεδομένα τους χρησιμοποιώντας αλγόριθμους ανωνυμοποίησης δεδομένων. Τροποποιεί προσωπικές και ευαίσθητες πληροφορίες, εξαλείφει κάθε παραβίαση απορρήτου δεδομένων και έκθεση ευαίσθητων πληροφοριών. Εγκατεστημένο, ρυθμισμένο και ενσωματωμένο τοπικά σε κάθε ηλεκτρονικό υπολογιστή, σε ερευνητικές επιχειρησιακές ροές εργασίας, επεξεργάζεται και αποθηκεύει αρχεία που προσαρμόζονται σε διαφορετικές ανάγκες και ροές δεδομένων. Χρησιμοποιεί αλγόριθμους k-ανωνυμίας και km-ανωνυμίας, δημογραφικά στατιστικά στοιχεία και , ψευδό - ανωνυμοποίηση, για την ανωνυμοποίηση των προσωπικών δεδομένων με την μικρότερη δυνατή απώλεια δεδομένων. (Amnesia software v. 1.3.2 (04/2023))

Σύμφωνα με τις πληροφορίες από τη συνέντευξη που μας παραχώρησε ο κ. Δημακόπουλος Νικόλαος για το λογισμικό Amnesia (συνέντευξη 10ος/2022), η ιδέα και το εργαλείο ανήκουν στον Δρ. Μανώλη Τερροβίτη ο οποίος είναι ερευνητής του Ερευνητικού Κέντρου Αθηνά και καθοδηγεί την ανάπτυξη και βελτίωση του εργαλείου. Ο κ. Δημακόπουλος είναι πτυχιούχος του Τμήματος Πληροφορικής κ' Τηλεπικοινωνιών του ΕΚΠΑ, κατέχει μεταπτυχιακό τίτλο στην Επιστήμη Δεδομένων και την Μηχανική Μάθηση από το ΕΜΠ. Ο ρόλος του κ. Δημακόπουλου στο Amnesia έχει να κάνει με την ανάπτυξη και την συντήρηση του λογισμικού στο προγραμματιστικό κομμάτι από το 2019. Επίσης παρέχει τεχνική υποστήριξη (support) σε χρήστες που έχουν κάποιο πρόβλημα με το λογισμικό.

Ερωτηθείς ο κ. Δημακόπουλος για ποιους χρήστες προορίζεται το λογισμικό Amnesia, μας απάντησε πως αυτή τη στιγμή το λογισμικό το χρησιμοποιούν πρωτίστως ερευνητές προκειμένου να ανωνυμοποιήσουν τα δεδομένα τους και να βγάλουν διάφορα στατιστικά

στοιχεία από αυτά. Αλλά ο δικός τους στόχος είναι να μπει και στην παραγωγή (industry) στον ιδιωτικό τομέα και ιδιαίτερα να το χρησιμοποιήσουν εταιρίες που ασχολούνται με την ασφάλεια (security) και αναλαμβάνουν την ασφάλεια σε τρίτες εταιρίες. Να το χρησιμοποιήσουν εκεί και να προσπαθήσουν να ανωνυμοποιήσουν τα δεδομένα τους είτε σε βάσεις δεδομένων είτε στα αρχεία που μπορεί να έχουν σε διάφορους υπολογιστές ούτως ώστε να μπορέσουν οι απλοί πολίτες να προστατευτούνε, γιατί σε μία ενδεχόμενη επίθεση από κακόβουλους επιτιθέμενους (hackers) μπορεί να βρεθούν εκτεθειμένοι. Συνεχίζοντας ο κ. Δημακόπουλος τόνισε πως συνήθως οι εταιρίες κάνουν πολύ απλή ανωνυμοποίηση, χρησιμοποιώντας μόνο όνομα – επίθετο και ΑΜΚΑ. Αφαιρούν αυτά και αφήνουν όλα τα υπόλοιπα. Έτσι όμως πέφτουν σε αυτήν την παγίδα των επιτιθέμενων. Μας ανέφερε το χαρακτηριστικό παράδειγμα με το βραβείο Netflix (βλ. ενότητα 2.1), όπου η εταιρία πριν κάποια χρόνια έβγαλε ένα διαγωνισμό και είπε θα βγάλω τα δεδομένα μου χωρίς να κοινοποιήσω το ονομ/μο και τον τόπο κατοικίας των χρηστών, θα αφήσω την ηλικία και τις βαθμολογίες που μπορεί να έχουνε βάλει στις διάφορες σειρές, προκειμένου να τα πάρουνε είτε ερευνητές είτε φοιτητές και να τρέξουμε μοντέλα μηχανικής μάθησης.

Υπήρξαν λοιπόν κάποιοι ερευνητές που πήραν αυτά τα δεδομένα, αυτά τα σύνολα δεδομένων (datasets) και με το συνδυασμό δεδομένων από imdb όπου καταχωρούνται οι βαθμολογίες των τηλεοπτικών προγραμμάτων (ταινίες και σειρές), κατάφεραν να βρούνε και το ονομ/μο και τον τόπο κατοικίας των χρηστών και να τους στοχοποιήσουνε. Οπότε, όπως καταλαβαίνετε, δεν υπάρχει τίποτε ασφαλέστερο από μία ανωνυμοποίηση – γενικοποίηση των δεδομένων, γιατί ουσιαστικά κάνουμε πιο γενικά τα δεδομένα, αυτό είναι το concept της ανωνυμοποίησης.

Το εργαλείο είναι γραμμένο σε java και είναι διαθέσιμο ως λογισμικό ανοιχτού κώδικα που εκτελείται σε όλα τα μεγάλα λειτουργικά συστήματα (Windows, Linux) και είναι ακόμα υπό ενεργή ανάπτυξη με τελευταία ενημέρωση την αναβάθμιση του 2022.

3.1 Έννοιες σχετικές με το λογισμικό

- Τα σύνολα δεδομένων που μπορεί να επεξεργαστεί το λογισμικό Amnesia αποθηκεύονται ως οριοθετημένα αρχεία κειμένου με κατάληξη .txt ή αρχεία excel με κατάληξη .csv . Τα ανωνυμοποιημένα αρχεία αποθηκεύονται στην ίδια μορφή με τα αρχικά αρχεία.

- Κάθε γραμμή στο αρχείο κειμένου είναι μια διαφορετική εγγραφή και κάθε ξεχωριστή τιμή της εγγραφής διαχωρίζεται από την επόμενη με έναν οριοθέτη. Ο χρήστης πρέπει να παρέχει τον οριοθέτη που χρησιμοποιείται στο αρχικό αρχείο, στο εργαλείο, κατά την εισαγωγή ενός συνόλου δεδομένων.

Τα σύνολα δεδομένων που υποστηρίζονται από το Amnesia είναι σχεσιακοί πίνακες, συλλογές συνόλων και πίνακες αντικειμενικής σχέσης.

- Οι σχεσιακοί πίνακες (relational tables), έχουν έναν σταθερό αριθμό στηλών και ως αποτέλεσμα, κάθε εγγραφή έχει τον ίδιο αριθμό τιμών. Κάθε στήλη μπορεί να έχει διαφορετικούς τύπους δεδομένων .

- Τα σύνολα δεδομένων (datasets), είναι σύνολα δεδομένων εγγραφών αυθαίρετου μήκους. Κάθε εγγραφή μπορεί να έχει έναν αυθαίρετο αριθμό τιμών του ίδιου τύπου δεδομένων (αυτή τη στιγμή υποστηρίζονται συμβολοσειρές).

- Οι αντικειμενο-σχεσιακοί πίνακες (object – relational tables), είναι ο συνδυασμός των παραπάνω. Αυτοί οι πίνακες έχουν έναν σταθερό αριθμό στηλών, αλλά μια στήλη είναι ένα σύνολο, δηλαδή περιέχει έναν αυθαίρετο αριθμό τιμών του ίδιου τύπου. Για να αναλύσει σωστά ένα οριοθετημένο αρχείο που περιέχει έναν πίνακα αντικειμενικής σχέσης, ο χρήστης πρέπει να παρέχει στο Amnesia δύο οριοθέτες: διαχωρισμό τιμών διαφορετικών στηλών και διαχωρισμό διαφορετικών τιμών στη στήλη συνόλου.

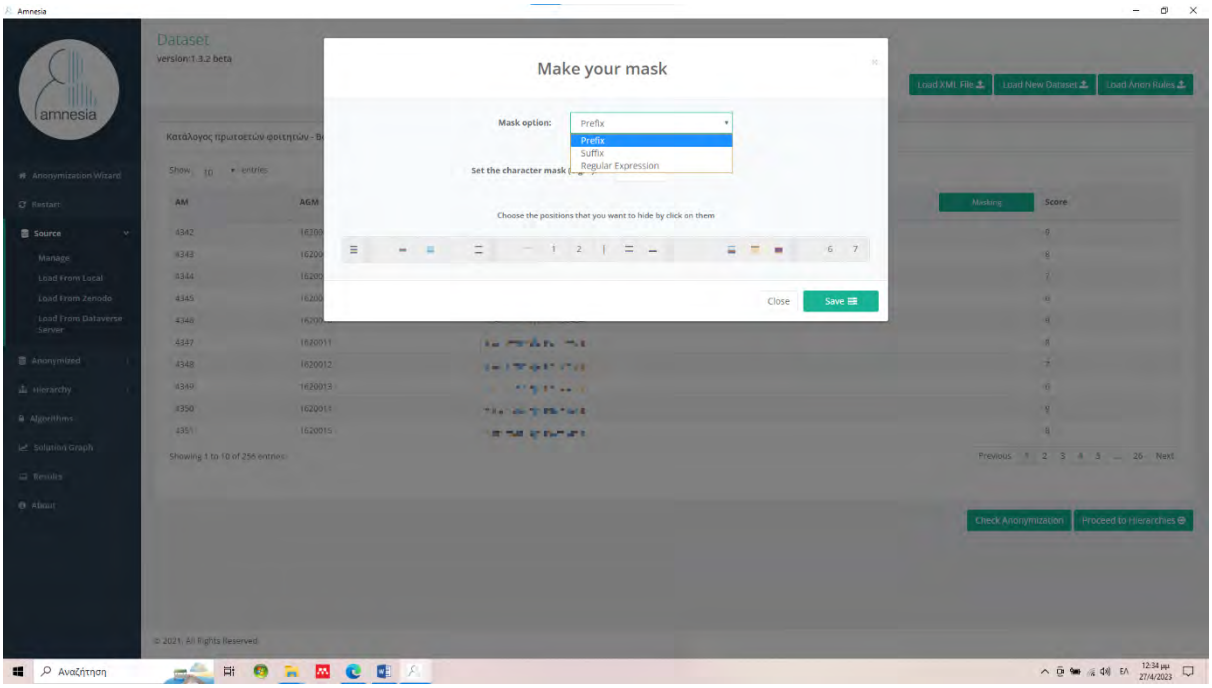
- Το Amnesia υποστηρίζει τέσσερις τύπους δεδομένων: συμβολοσειρές, ακέραιους, διπλούς (κινητή υποδιαστολή) και ημερομηνίες. Κατά τη φόρτωση ενός συνόλου δεδομένων, το Amnesia θα προσπαθήσει να μαντέψει τον τύπο δεδομένων, αλλά θα το κάνει με βάση μόνο τις πρώτες γραμμές των εισαγόμενων δεδομένων. Ο χρήστης θα πρέπει να ελέγξει τα δεδομένα της κάθε στήλης και να επέμβει διορθωτικά αν χρειάζεται.

- Η συγκάλυψη (βλ. εν. 1.9) που μπορούμε να εφαρμόσουμε μέσω του λογισμικού, πχ σε τηλεφωνικό αριθμό ή σε διεύθυνση ηλεκτρονικού ταχυδρομείου, έχει τις εξής επιλογές:

    - Επιλογή προθέματος (prefix) μέσω της οποίας επιλέγουμε τον χαρακτήρα συγκάλυψης και το εύρος των γραμμάτων ή αριθμών που επιθυμούμε να συγκαλύψουμε

- Επιλογή κατάληξης (suffix) μέσω της οποίας επιλέγουμε τον χαρακτήρα συγκάλυψης και το εύρος των γραμμάτων ή αριθμών που επιθυμούμε να συγκαλύψουμε στο πρώτο τμήμα μιας διεύθυνσης ηλεκτρονικού ταχυδρομείου, δηλαδή πριν το @.

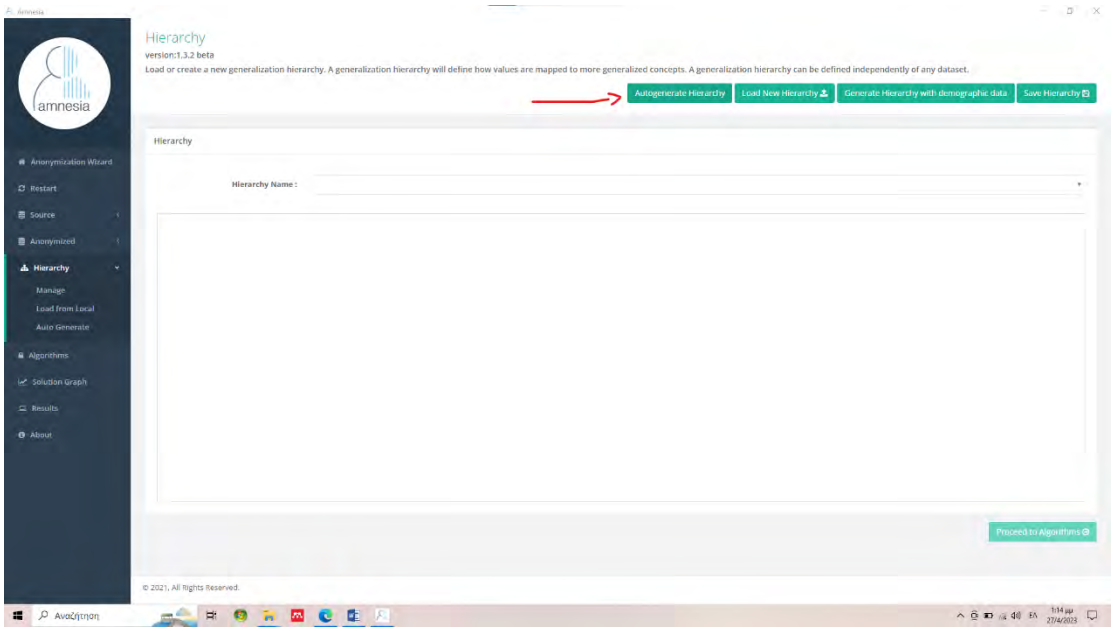
- Επιλογή κοινής έκφρασης (regular expression) μέσω της οποίας επιλέγουμε συνδυασμό γραμμάτων ή αριθμών ή και των δύο προκειμένου να συγκαλύψουμε από τα δεδομένα μας πχ όλους τους τηλεφωνικούς κωδικούς πόλεων (επειδή έχουν κοινή έκφραση πχ 24210 κλπ)



Εικόνα 9: Εισαγωγή συγκάλυψης με το λογισμικό Amnesia.

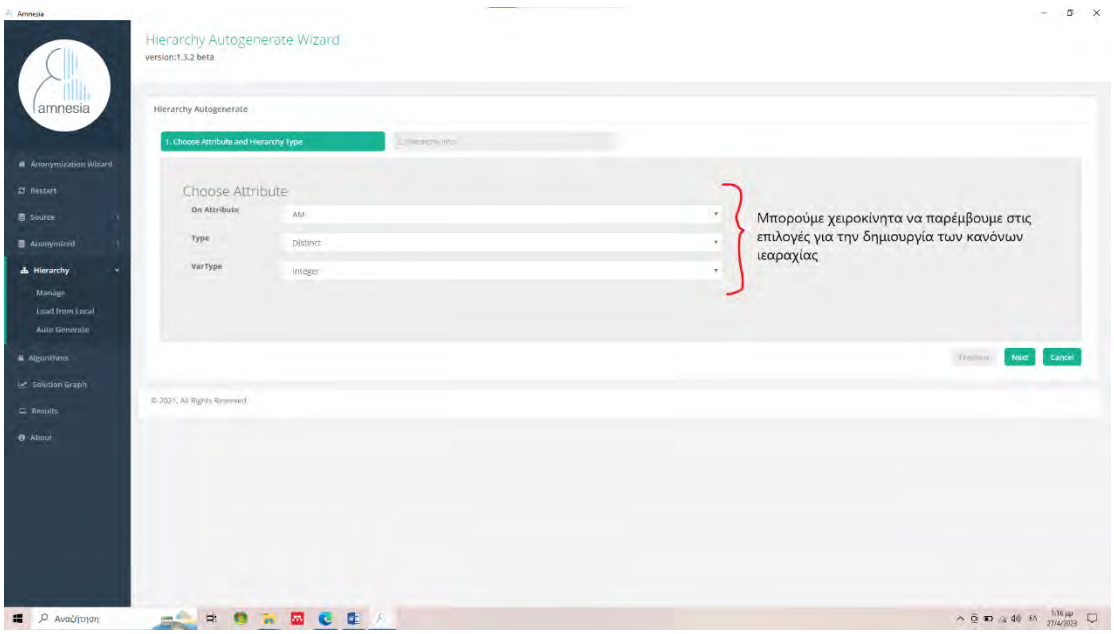
- Οι ιεραρχίες γενίκευσης (βλ. εν. 1.4) εφαρμόζονται στα σύνολα δεδομένων με βάση ορισμένους κανόνες, που ονομάζονται κανόνες ιεραρχίας. Αυτοί οι κανόνες μπορούν να δημιουργηθούν είτε αυτόματα μέσω της αντίστοιχης επιλογής από το λογισμικό, είτε να επέμβει ο χρήστης δημιουργώντας τους χειροκίνητα σε συγκεκριμένα σημεία.

Για την αυτόματη επιλογή χρησιμοποιούμε την ένδειξη Autogenerate:



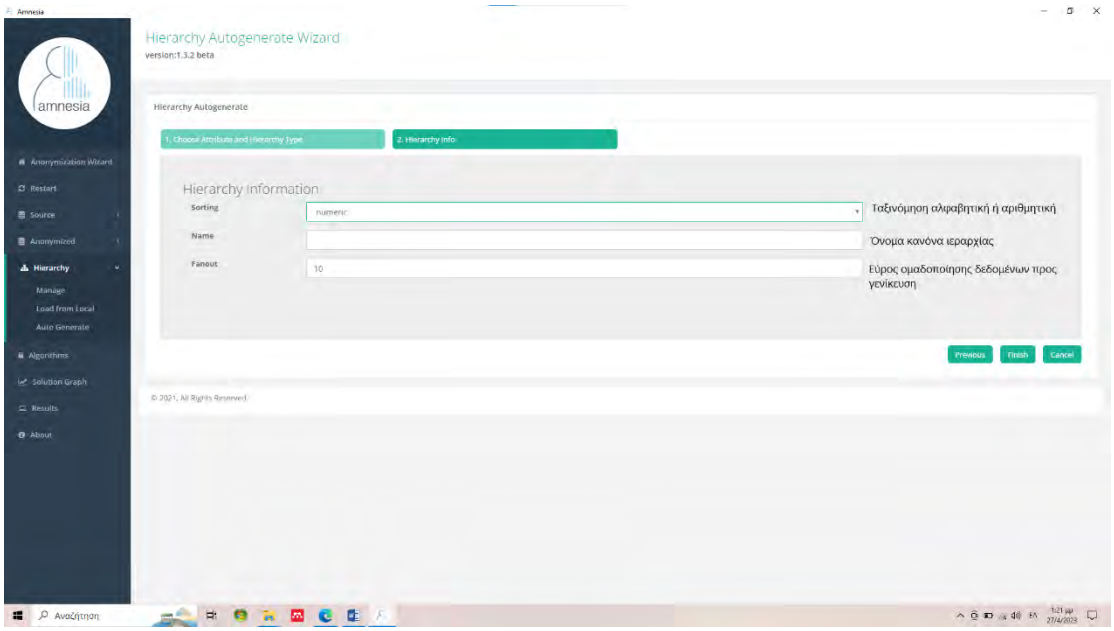
Εικόνα 10: Εισαγωγή Ιεραρχιών γενίκευσης με το λογισμικό Amnesia.

Αφού ενεργοποιηθεί το μενού έχουμε τη δυνατότητα να αφήσουμε τις αυτόματες επιλογές του λογισμικού ή να επέμβουμε χειροκίνητα και να αλλάξουμε το είδος, τον τύπο και τις ιδιότητες του συνόλου δεδομένων που έχουμε επιλέξει ώστε να δημιουργήσουμε κανόνες ιεράρχησης



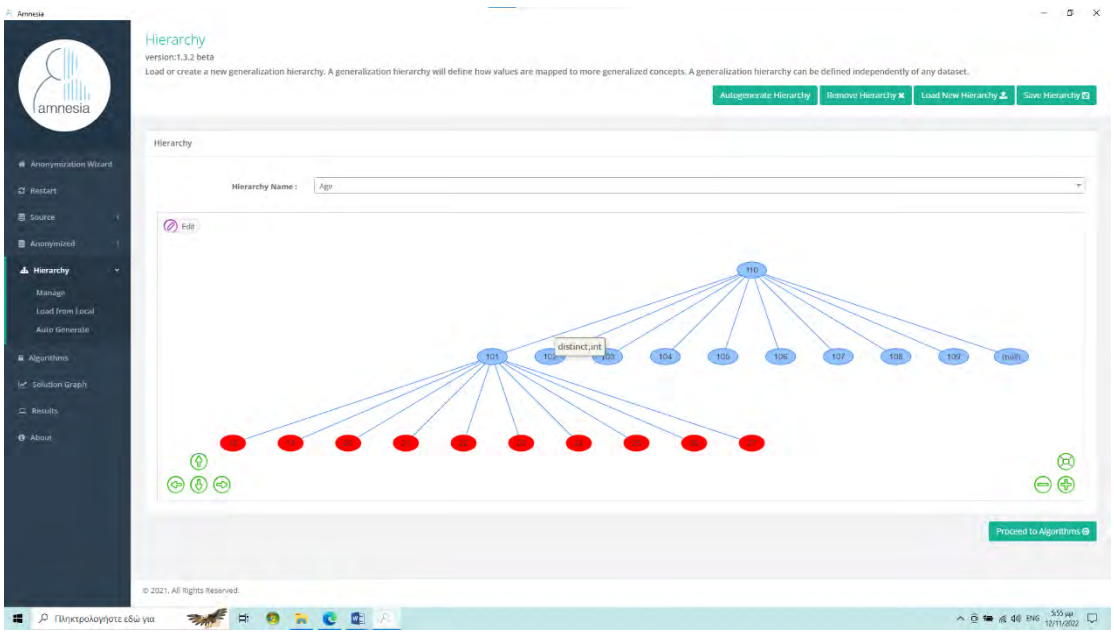
Εικόνα 11: Ιεραρχία γενίκευσης - επιλογή τύπου.

Το τελικό βήμα για τη δημιουργία των κανόνων αποτελείται από το πως επιθυμεί ο χρήστης την ταξινόμηση των δεδομένων του (sorting), το όνομα των κανόνων (name) που θα εφαρμόσει και το εύρος (μέγιστος αριθμός) αρχικών δεδομένων που θα χρησιμοποιηθούν για το κάθε επίπεδο γενίκευσης.

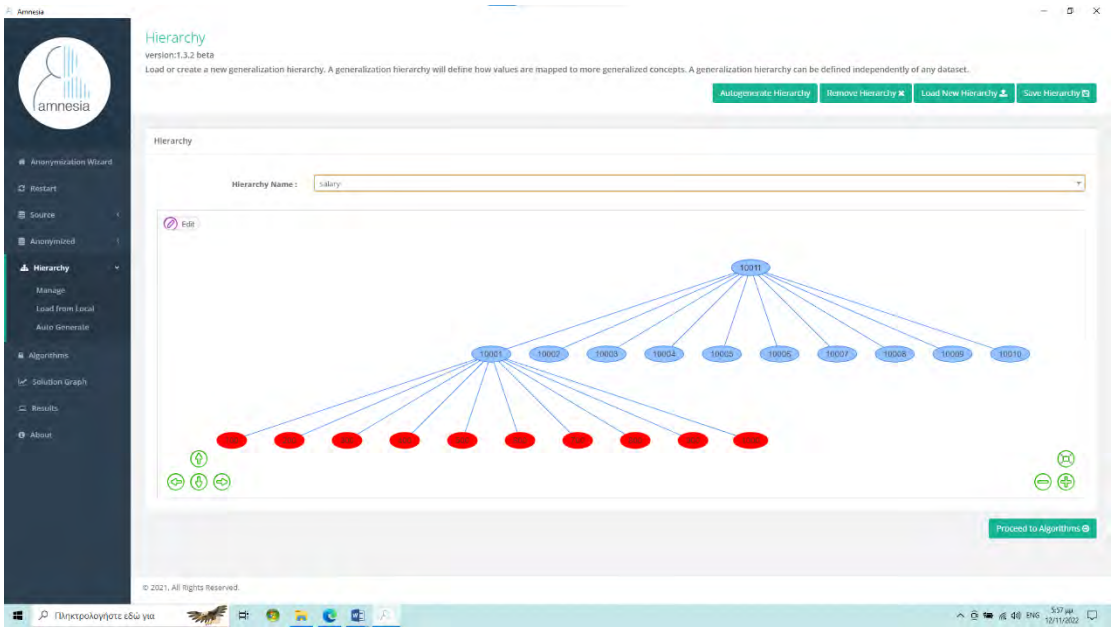


Εικόνα 12: Ιεραρχία γενίκευσης, ταξινόμηση και εύρος.

Αυτό γραφικά παρουσιάζεται ως κόκκινα και μπλε σύννεφα. Το πρώτο επίπεδο περιλαμβάνει κόκκινα σύννεφα με εύρος ομαδοποίησης αυτό που επιλέξαμε (στο fanout) και μπλε σύννεφα είναι το κάθε επίπεδο γενίκευσης. Γι' αυτό επειδή μοιάζει με δέντρο ονομάζουμε τα κόκκινα σύννεφα κλαδιά, τα μπλε σύννεφα κορμό και τα κορυφαία ρίζα.

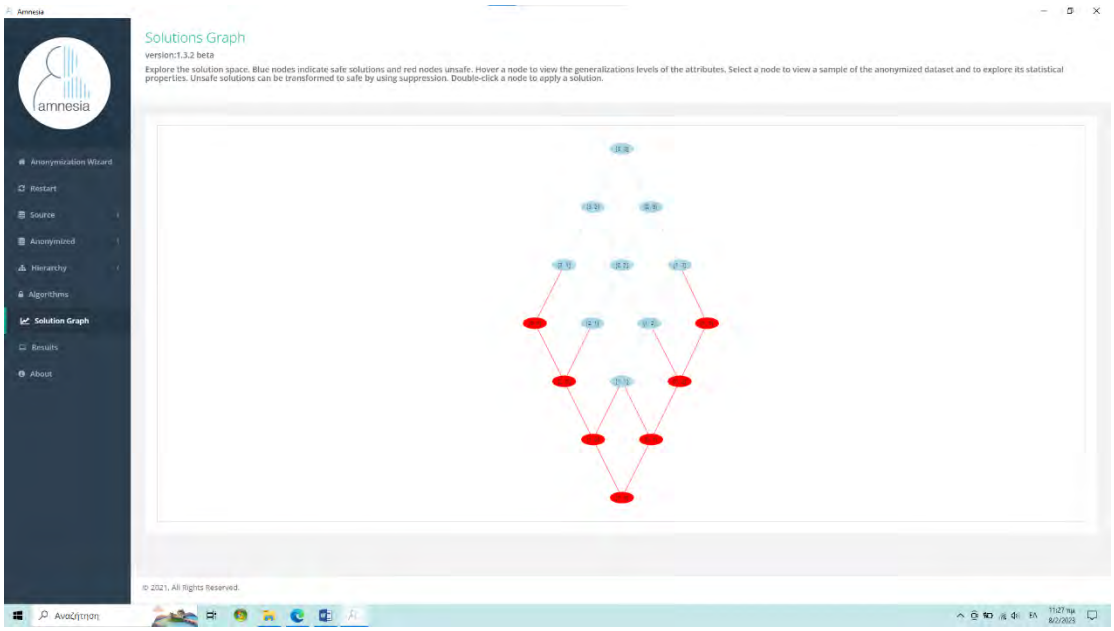


Εικόνα 13: Απεικόνιση ρίζας του δέντρου με το λογισμικό Amnesia. Εικόνα α



Εικόνα 14: Απεικόνιση ρίζας του δέντρου με το λογισμικό Amnesia. Εικόνα β

- Χώρος λύσεων (solution graph) είναι το τελευταίο στάδιο του λογισμικού. Σε αυτό το γραφικό περιβάλλον, μας παρουσιάζεται κάθε συνδυασμός δεδομένων (χαρακτηριστικών) και κανόνων ιεραρχίας με εφαρμογές ανωνυμοποίησης και στο οποίο παρουσιάζονται όλες οι ασφαλείς (μπλε κόμβοι) και οι μη ασφαλείς (κόκκινοι κόμβοι) λύσεις ανωνυμοποίησης. Τοποθετώντας τον κέρσορα σε κάθε ένα από τους κόμβους, ενημερωνόμαστε για το επίπεδο γενίκευσης που έχει πραγματοποιηθεί σε κάθε κόμβο.

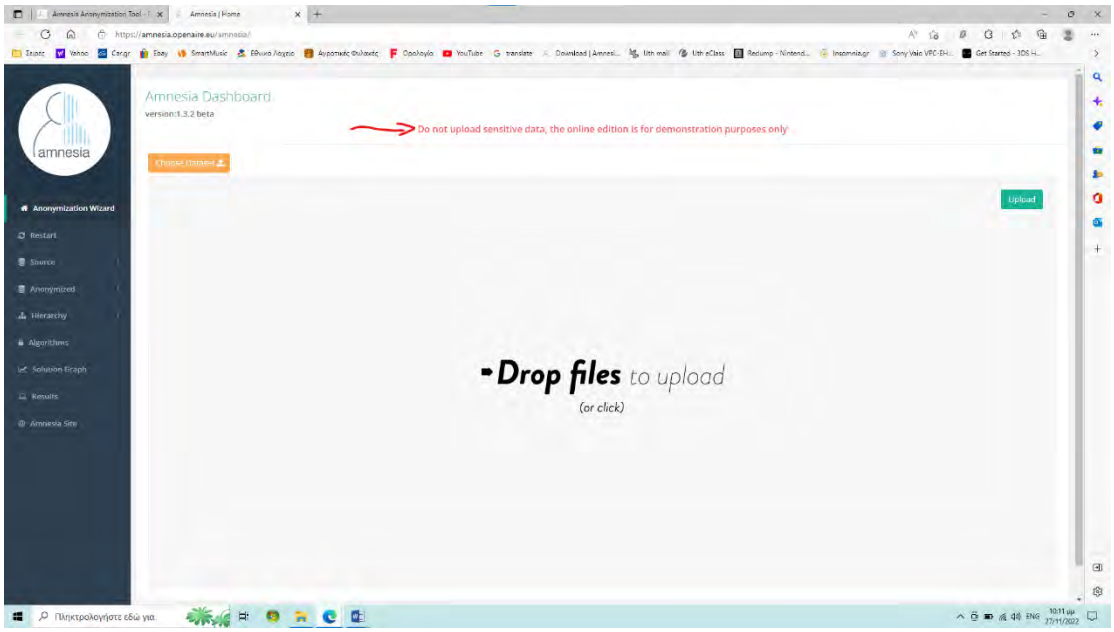


Εικόνα 15: Χώρος λύσεων με το λογισμικό Amnesia.

3.2 Οδηγίες εγκατάστασης

Το Amnesia μπορεί να χρησιμοποιηθεί είτε μέσω της διαδικτυακής του πλατφόρμας (online) είτε με εγκατάσταση σε τοπικό υπολογιστή.

1) Μέσω της διεύθυνσης https://amnesia.openaire.eu/index.html), (Amnesia software v. 1.3.2 (04/2023)), πάνω δεξιά κάνουμε κλικ στο κουμπί ONLINE DEMO και οδηγούμαστε έπειτα από τις συμβουλές - επισημάνσεις για την απαραίτητη νομοθεσία, στην αρχική οθόνη του λογισμικού όπου ενημερωνόμαστε ότι η online έκδοση είναι μόνο για επίδειξη του λογισμικού

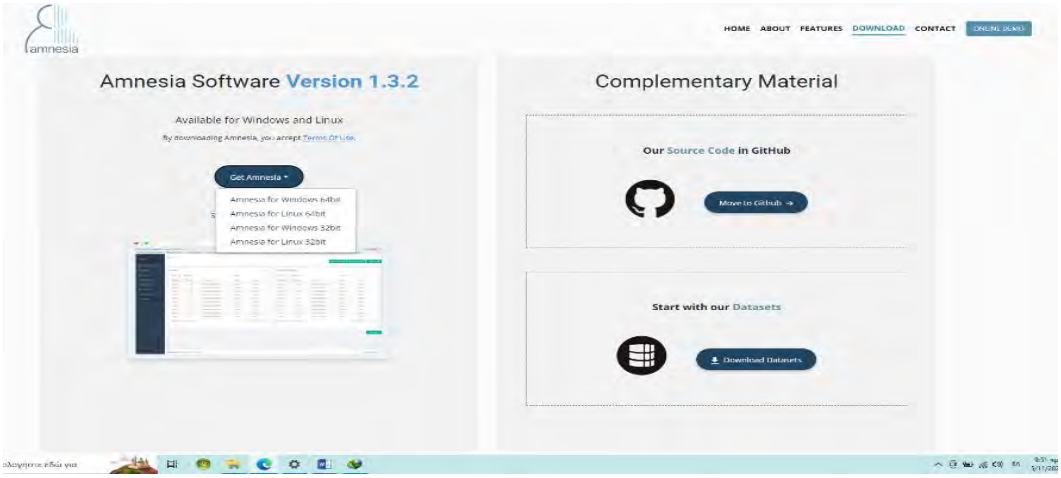


Εικόνα 16: On-line Έκδοση.

2) Μέσω της ιστοσελίδας του Amnesia, ακολουθώντας την επιλογή DOWNLOAD

ή Download | Amnesia Anonymization Tool - Data anonymization made easy (https://amnesia.openaire.eu/download.html), οδηγούμαστε στον τομέα όπου μπορούμε να κατεβάσουμε την εφαρμογή για περαιτέρω εγκατάσταση και χρήση στον υπολογιστή μας. Επιλέγουμε την ένδειξη «Get Amnesia», μας προβάλλονται τέσσερις επιλογές:

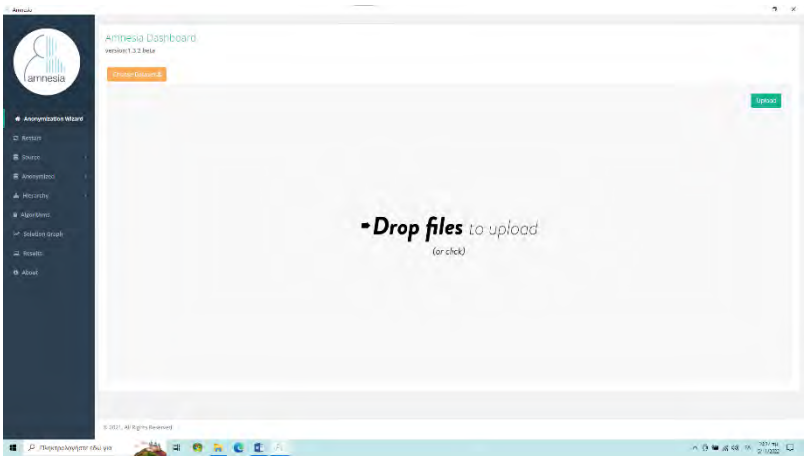
1. Amnesia για Windows 64Bit
2. Amnesia για Linux 64Bit
3. Amnesia για Windows 32Bit
4. Amnesia για Linux 32Bit



Εικόνα 17: Εγκατάσταση του Amnesia.

και επιλέγουμε την κατάλληλη σύμφωνα με το λειτουργικό πρόγραμμα που χρησιμοποιούμε στον υπολογιστή μας. Η εγκατάσταση του amnesia είναι μία τυπική εγκατάσταση λογισμικού.

Ξεκινώντας για πρώτη φορά την κεντρική εικόνα του λογισμικού Amnesia που όπως θα παρατηρήσουμε είναι όμοια με αυτή της δοκιμής που μας παρέχεται μέσω της επιλογής ONLINE DEMO στην ιστοσελίδα του λογισμικού.



Εικόνα 18: Αρχική εικόνα του λογισμικού.

3.3 Μέθοδος Συγκάλυψης (Masking)

Για τις ανάγκες αυτής της έρευνας θα χρησιμοποιήσουμε ένα σύνολο δεδομένων EXCEL που αφορά βαθμολογίες πρωτοετών φοιτητών του πανεπιστημίου Θεσσαλίας. Το σύνολο

δεδομένων περιλαμβάνει τα εξής: Αριθμός Μητρώου (ΑΜ), Αριθμός Γενικού Μητρώου (ΑΓΜ), ονομ/μο, πατρώνυμο, email, Τηλ. Επικοινωνίας και βαθμολογίες.

Ο σκοπός μας είναι να μπορέσουμε να ανωνυμοποιήσουμε το σύνολο αλλά να παραμείνουν οι βαθμολογίες ορατές για έλεγχο.

Αρχικά αντιμετωπίζουμε δύο εμπόδια:

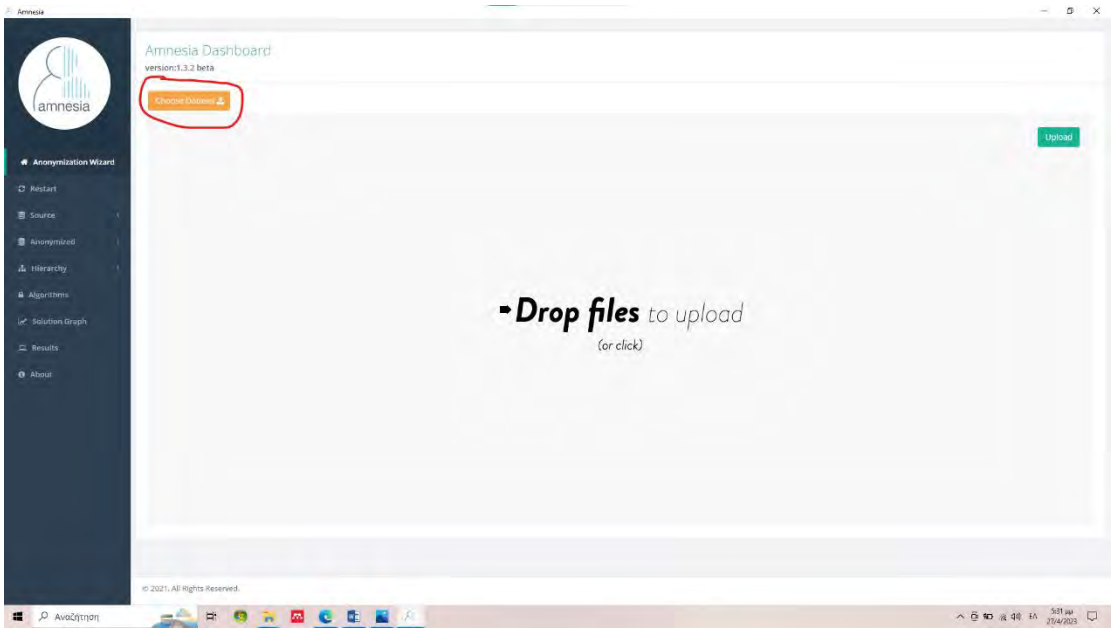
Α) το πρώτο είναι ότι η εφαρμογή Amnesia δεν αναγνώρισε τα αρχεία .xls. Για να μπορέσουμε να τα χρησιμοποιήσουμε, τα μετατρέψαμε σε αρχεία .csv μέσω του MS Excel και τα αποθηκεύσαμε εκ νέου ως ένα σύνολο δεδομένων με κατάληξη .csv.

Β) το δεύτερο εμπόδιο που αντιμετωπίσαμε ήταν ότι το λογισμικό δεν αναγνώρισε την ελληνική γλώσσα. Για να μπορέσουμε να χρησιμοποιήσουμε το σύνολο δεδομένων για το παράδειγμά μας έπρεπε να μην συμπεριλάβουμε το ονομ/μο των φοιτητών και το πατρώνυμό τους που ήταν στα ελληνικά ή να τα μεταφράσουμε στην αγγλική γλώσσα.

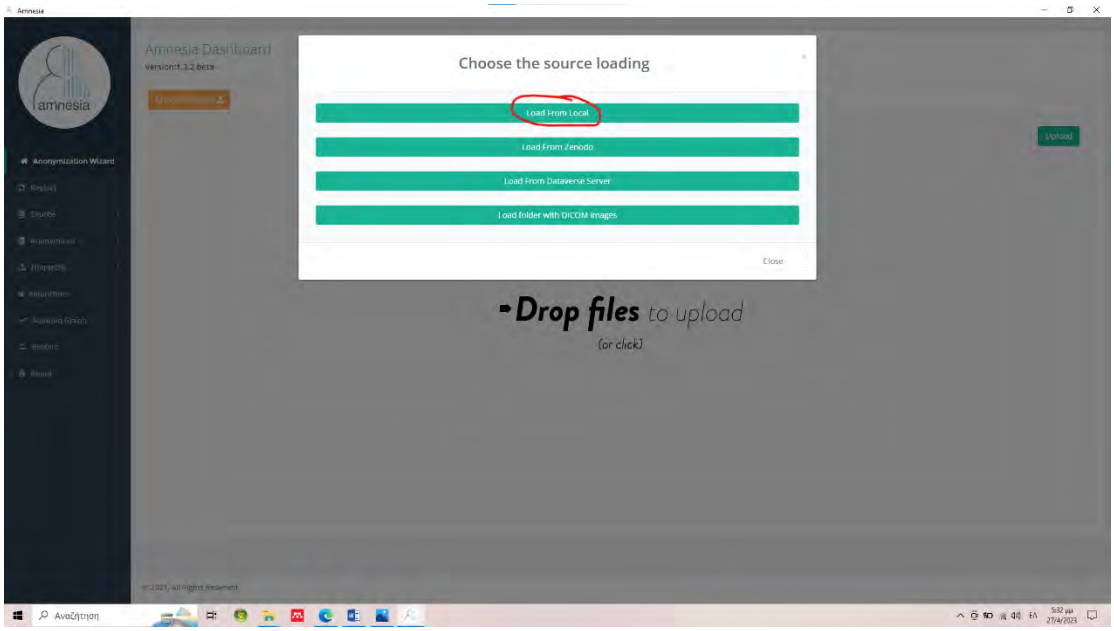
Ακολουθώντας το πλαίσιο του ΓΚΠΔ θα ανωνυμοποιήσουμε τα δεδομένα από τον αριθμό μητρώου (ΑΜ), τον αριθμό γενικού μητρώου (AGM), την διεύθυνση ηλεκτρονικού ταχυδρομείου (email), τον αριθμό τηλεφώνου (Phone) και τις βαθμολογίες (Grades).

Αρχεία συνόλου δεδομένων μπορούμε να φορτώσουμε στο λογισμικό είτε στη δοκιμαστική έκδοση (online demo) είτε στον τοπικό μας δίσκο. Επιλέγουμε τη δεύτερη μέθοδο εισαγωγής.

Από την επιλογή Choose Dataset →Load From Local επιλέγουμε το αρχείο με κατάληξη .csv του παραδείγματός μας.

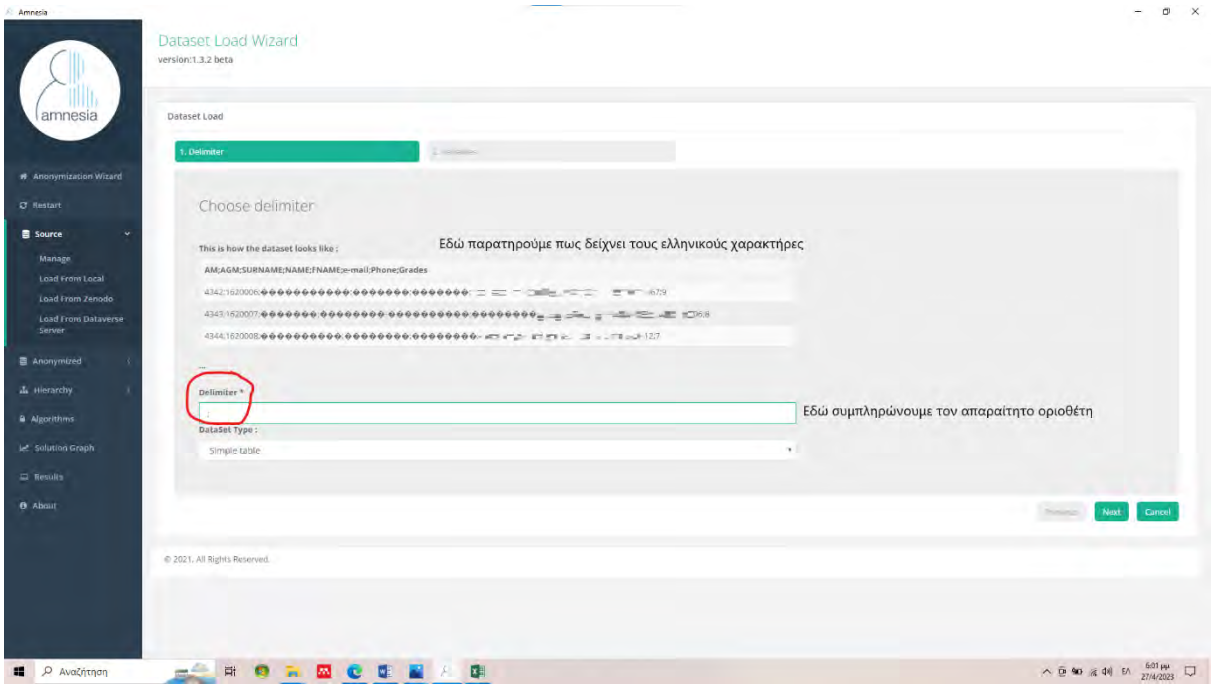


Εικόνα 19: Διαδικασία συγκάλυψης. Εισαγωγή συνόλου δεδομένων. Εικόνα α

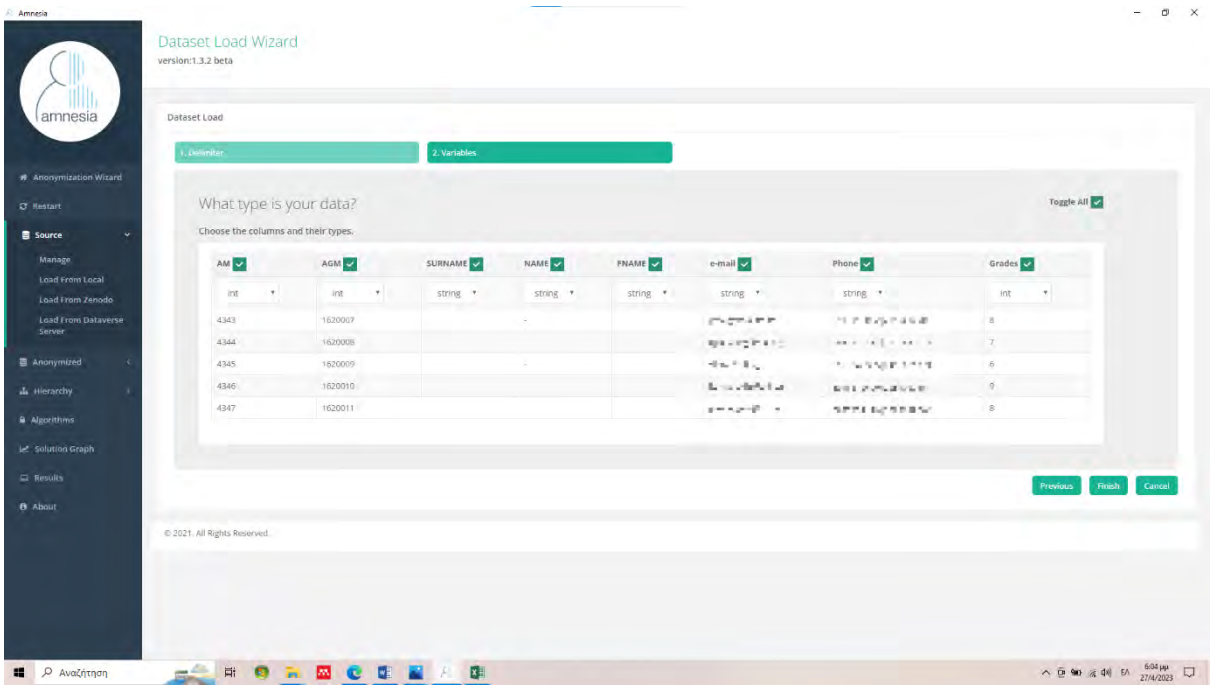


Εικόνα 20: Διαδικασία συγκάλυψης. Εισαγωγή συνόλου δεδομένων. Εικόνα β

Παρατηρούμε πως υποστηρίζει την ελληνική γραμματοσειρά για τις μεταβλητές (στήλες δεδομένων) που αφορούν όνομα, επώνυμο και πατρώνυμο. Για να μπορέσουμε να προχωρήσουμε πρέπει να δηλώσουμε τον οριοθέτη με βάση τον οποίο το λογισμικό θα μπορέσει να διαχωρίσει τις στήλες του συνόλου των δεδομένων.



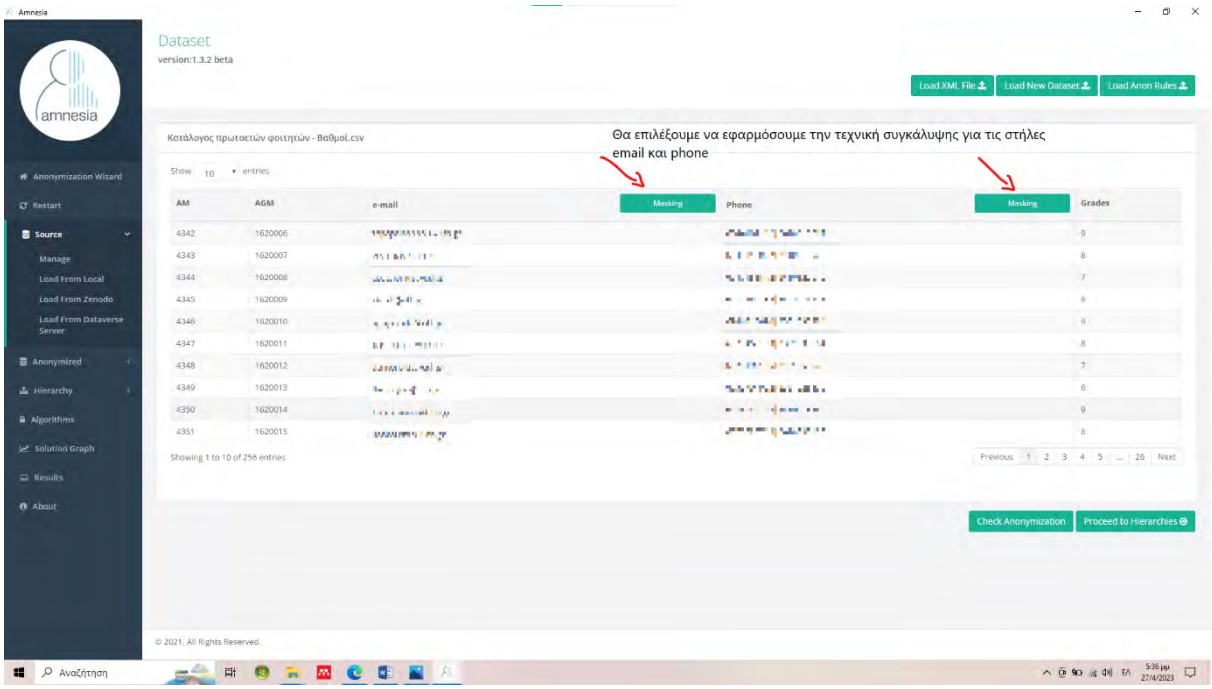
Εικόνα 21: Διαδικασία συγκάλυψης. Επιλογή οριοθέτη και χαρακτηριστικών. Εικόνα α



Εικόνα 22: Διαδικασία συγκάλυψης. Επιλογή οριοθέτη και χαρακτηριστικών. Εικόνα β

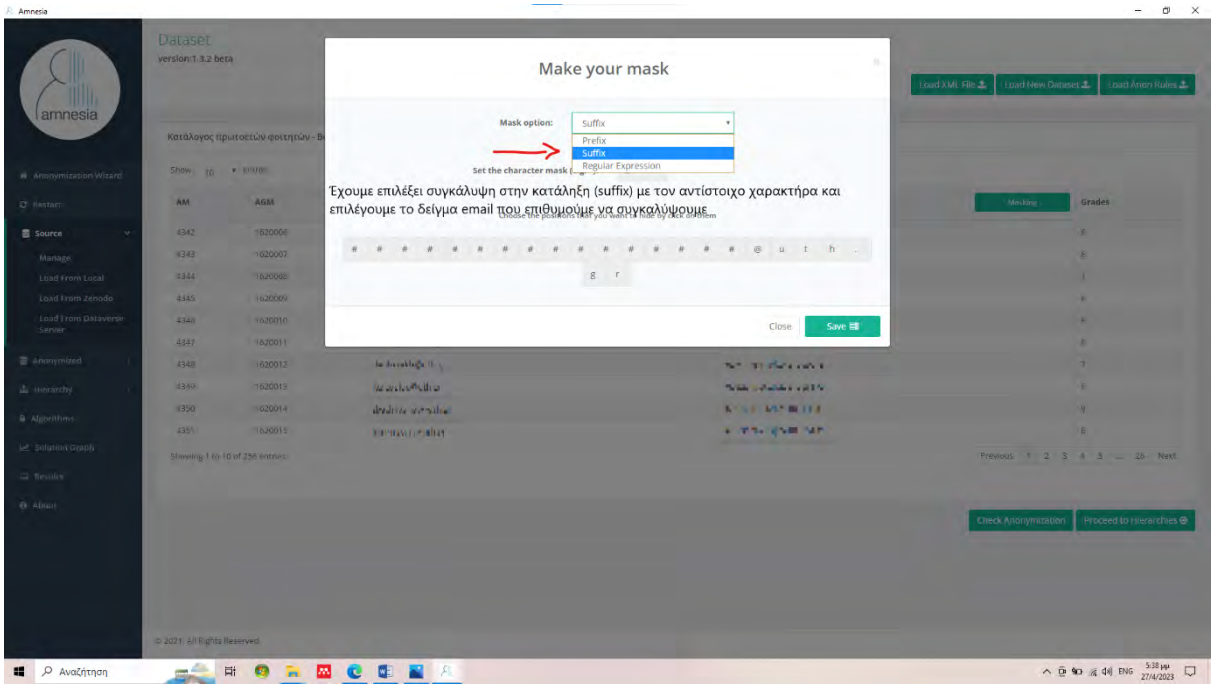
Από-επιλέγουμε αυτές τις κατηγορίες, δηλαδή τις SURNAME, NAME και FNAME και στο σύνολο δεδομένων θα εφαρμόσουμε την συγκάλυψη (επιλογή Masking) στην κατηγορία του mail και phone, διότι είναι μια τεχνική που μας επιτρέπει να αποκρύψουμε χαρακτηριστικά

χωρίς να χρειαστεί να τα παραλείψουμε ή να τα γενικεύσουμε. Επίσης θα αποτελέσει την βάση με την οποία στο τέλος θα μπορέσει ο κάθε ενδιαφερόμενος φοιτητής να ελέγξει την βαθμολογία του χωρίς να διαρρεύσουν τα προσωπικά του δεδομένα.



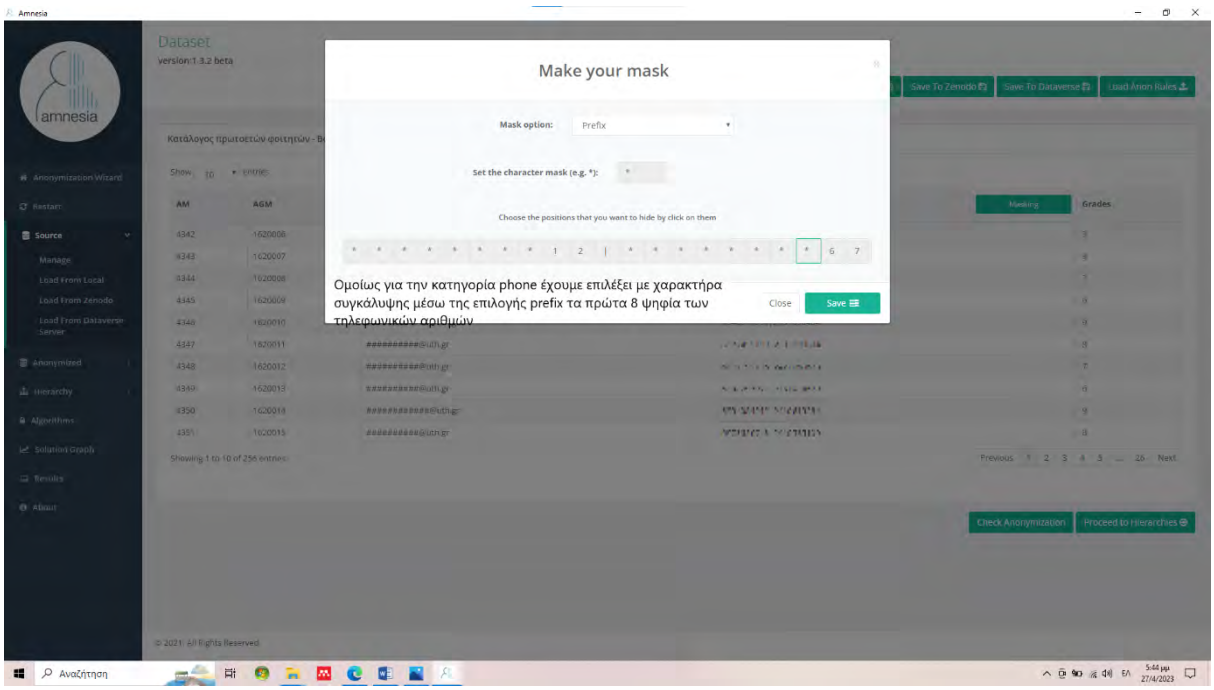
Εικόνα 23: Εφαρμογή συγκάλυψης σύμφωνα με το χρήστη.

Για την κατηγορία email χρησιμοποιούμε την επιλογή Suffix (βλ. σελ. 37 του παρόντος), τον χαρακτήρα της επιλογής μας και το εύρος που επιθυμούμε να συγκαλύψουμε.



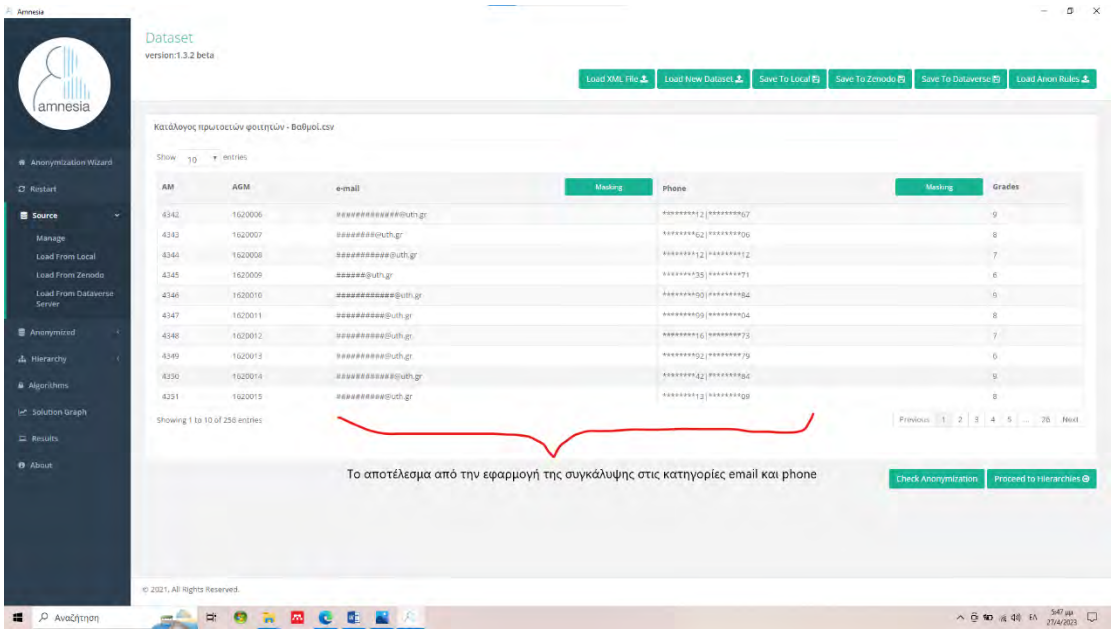
Εικόνα 24: Εφαρμογή συγκάλυψης σε email.

Αντίστοιχα για την κατηγορία phone χρησιμοποιούμε την επιλογή Prefix (βλ. σελ. 36 του παρόντος), τον χαρακτήρα της επιλογής μας και το εύρος που επιθυμούμε να συγκαλύψουμε.



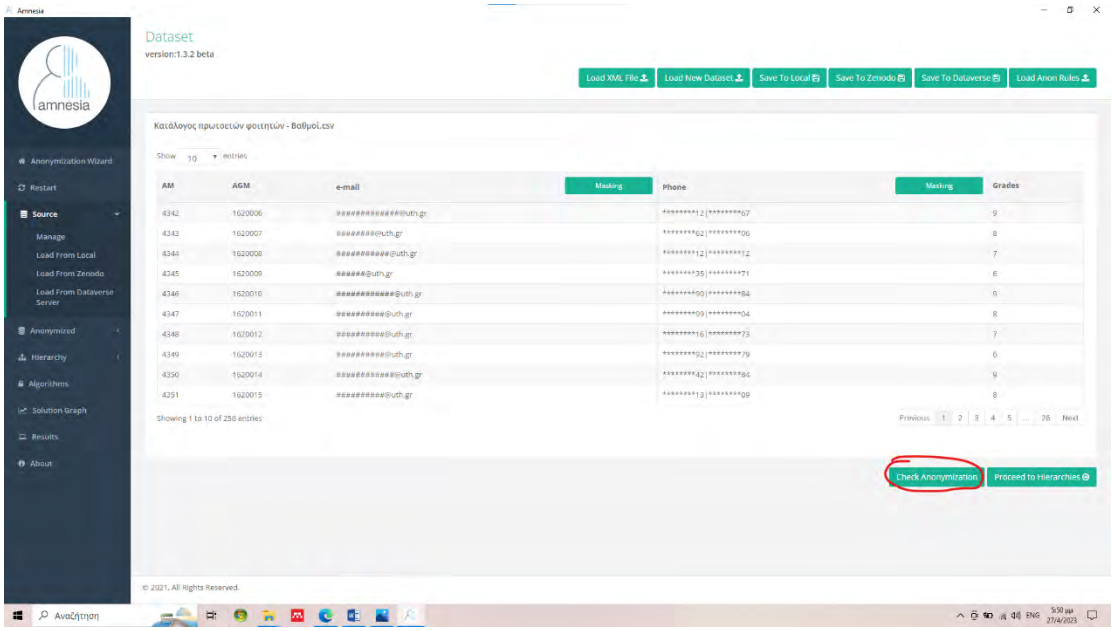
Εικόνα 25: Εφαρμογή συγκάλυψης σε phone.

Αφού αποθηκεύσουμε τις επιλογές μας στο επόμενο παράθυρο βλέπουμε το αποτέλεσμα από την εφαρμογή της συγκάλυψης στις αντίστοιχες κατηγορίες.

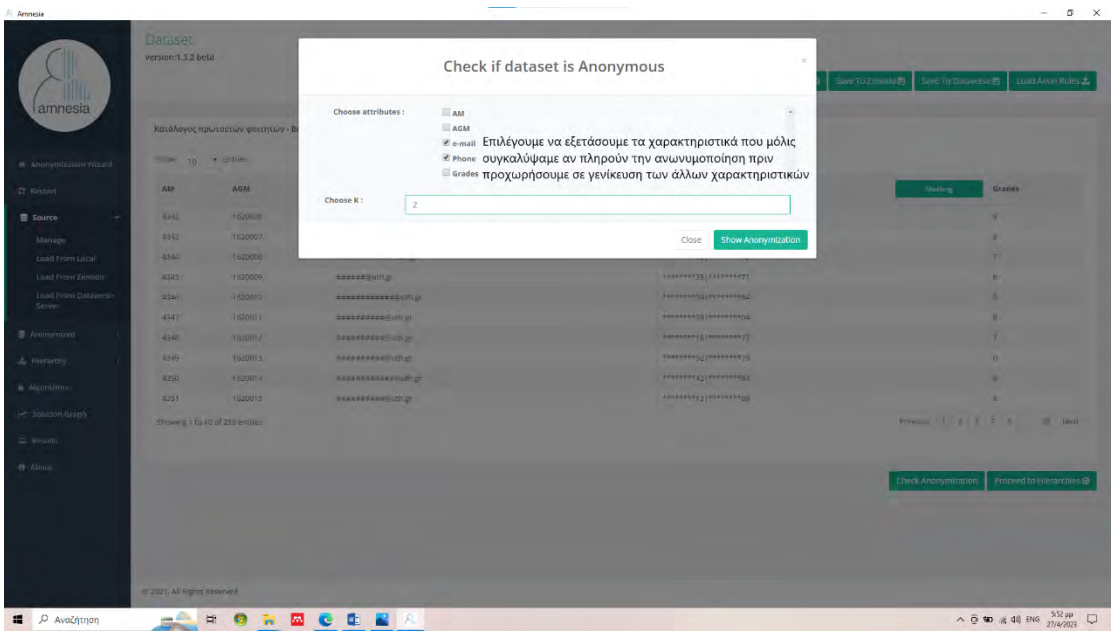


Εικόνα 26: Αποτέλεσμα συγκάλυψης

Με την επιλογή Check Anonymization μας δίνεται η δυνατότητα να ελέγξουμε στο σημείο που βρισκόμαστε αν έχουμε επιτύχει την επιθυμητή ανωνυμοποίηση. Από τις διαθέσιμες κατηγορίες επιλέγουμε αυτές στις οποίες έχουμε εφαρμόσει την συγκάλυψη, καθώς και τις βαθμολογίες των μαθητών, δηλαδή τα στοιχεία που θα δημοσιοποιήσουμε.

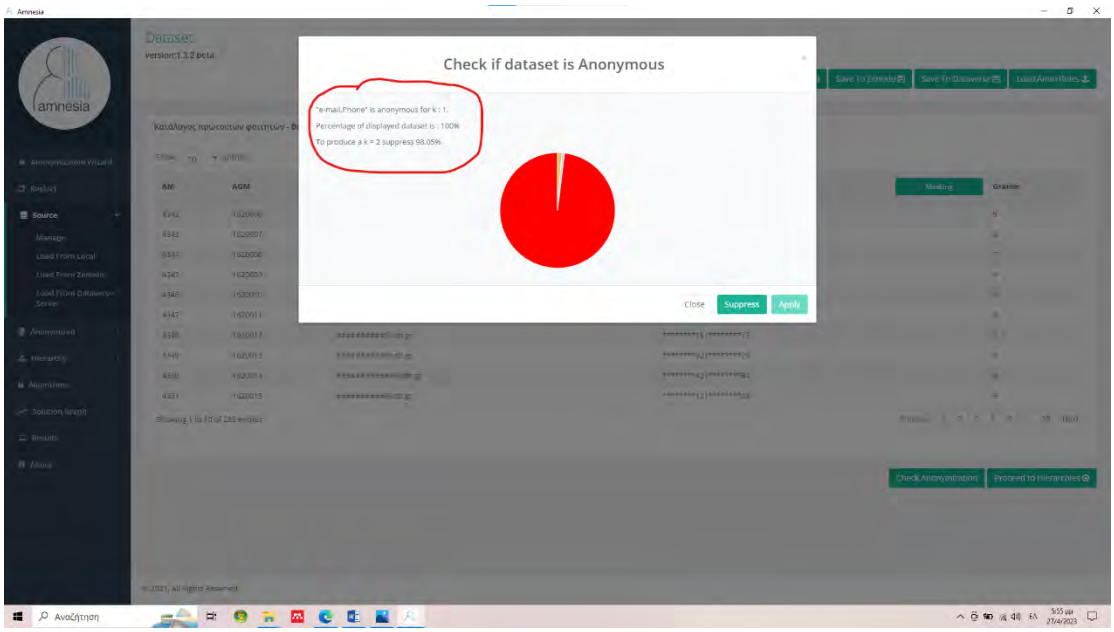


Εικόνα 27: Έλεγχος ανωνυμοποίησης συγκάλυψης. Εικόνα α



Εικόνα 28: Έλεγχος ανωνυμοποίησης συγκάλυψης. Εικόνα β

Βλέπουμε πως το λογισμικό μας ενημερώνει ότι το σύνολο δεδομένων είναι ανώνυμο αν αποφασίσουμε να χρησιμοποιήσουμε μόνο αυτά τα χαρακτηριστικά. Σε αυτό το σημείο πρέπει να θυμηθούμε την ολοκληρωτική απόκρυψη του χαρακτηριστικού (suppression). Όπου ουσιαστικά αναφερόμαστε στην εξ ολοκλήρου αφαίρεση ενός χαρακτηριστικού που δεν θέλουμε να συμπεριλάβουμε ή είναι αδιάφορο για το σκοπό που μελετάται το σύνολο δεδομένων μας (βλ. εν. 1.6, σελ.14). Επειδή όμως υπάρχουν και άλλα χαρακτηριστικά στο σύνολο δεδομένων μας, κλείνουμε αυτό το παράθυρο και επιλέγουμε την ένδειξη Proceed to Hierarchies ώστε να τα ενσωματώσουμε στην διαδικασία της ανωνυμοποίησης.

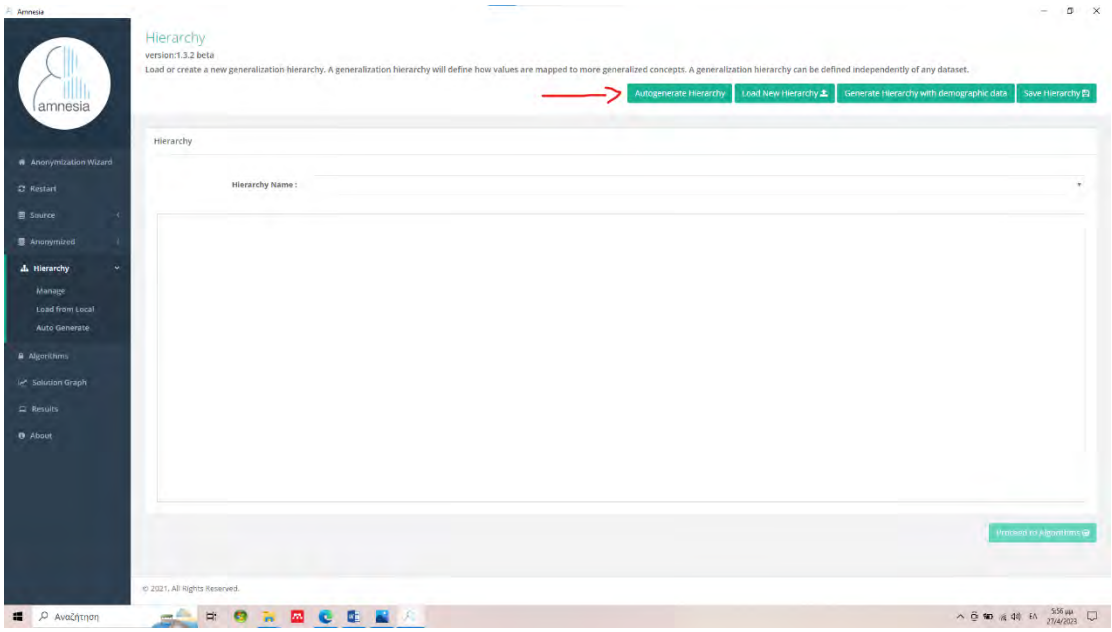


Εικόνα 29: Έλεγχος ανωνυμοποίησης για τα επιλεχθέντα χαρακτηριστικά.

3.4 Δημιουργία κανόνων ιεραρχίας

3.4.1 Εφαρμογή σε ποσοτικές μεταβλητές

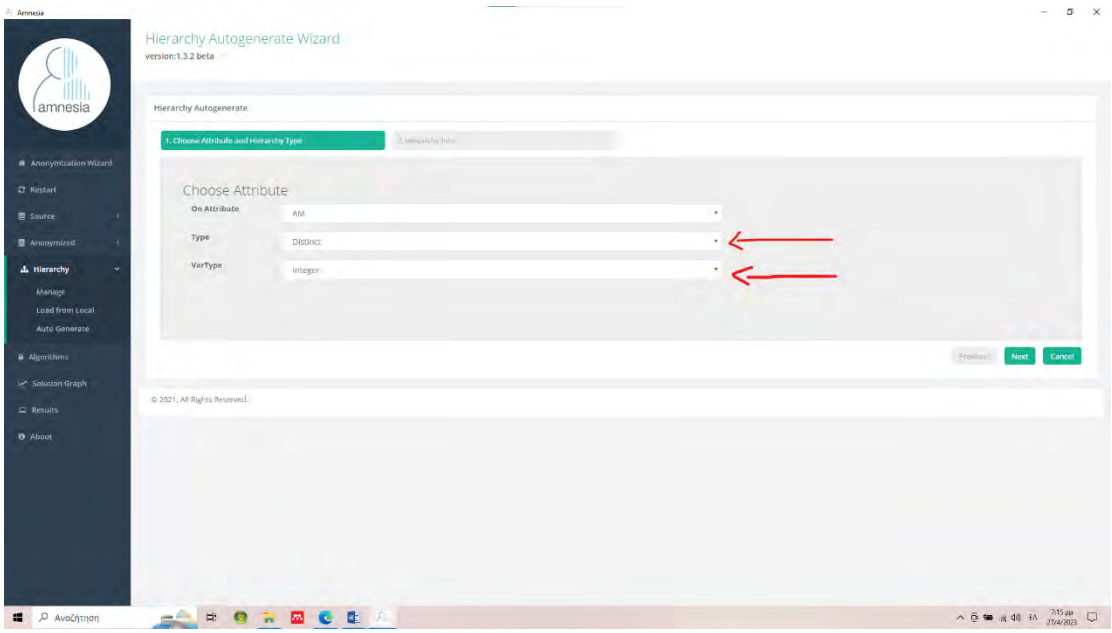
Στη συνέχεια θα δημιουργήσουμε κανόνες ιεραρχίας (βλ. εν. 1.4). Με τη επιλογή Autogenerate Hierarchies επιτρέπουμε στο Amnesia να ξεκινήσει αυτόματα την δημιουργία κανόνων ιεραρχίας και από τη λίστα των χαρακτηριστικών (attributes) επιλέγουμε διαδοχικά όσα θέλουμε να ρυθμίσουμε χειροκίνητα. Η αποθήκευση γίνεται στη θέση που επιθυμεί ο χρήστης σε μορφή αρχείου .txt.



Εικόνα 30: Εισαγωγή κανόνων ιεραρχίας.

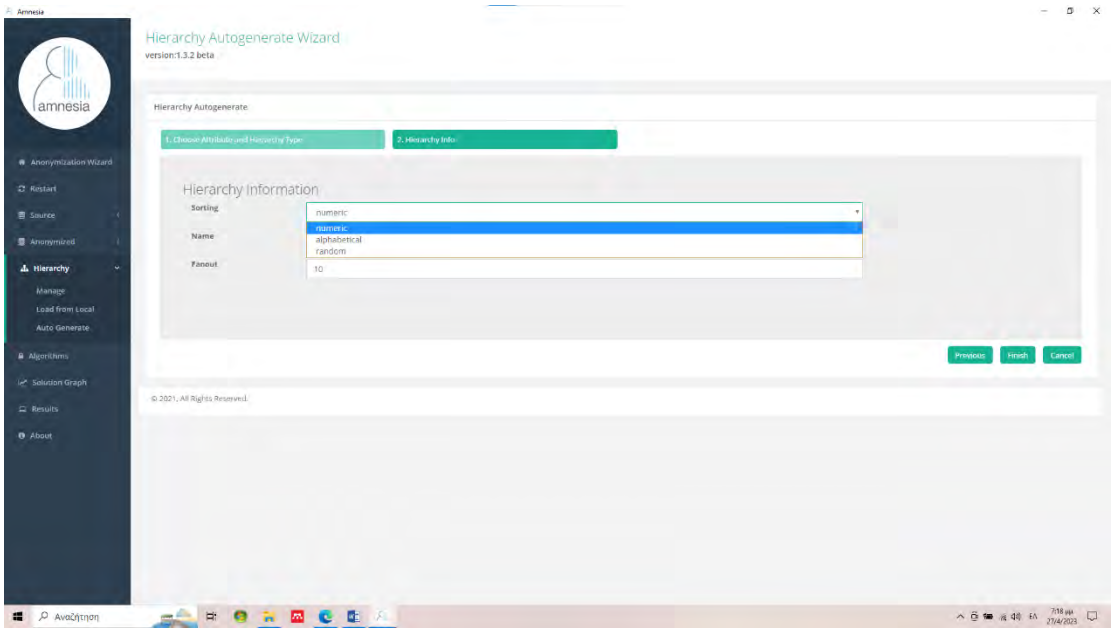
Επιλέγουμε το χαρακτηριστικό AM και τον τύπο εμφάνισης του χαρακτηριστικού:

- να εμφανίζεται διακριτά κάθε ένας αριθμός στην ομάδα (Distinct) και ο μεταβλητός τους τύπος (VarType) να είναι ακέραιος αριθμός καθώς πρόκειται για αριθμούς μητρώου. Επιλέγουμε Next

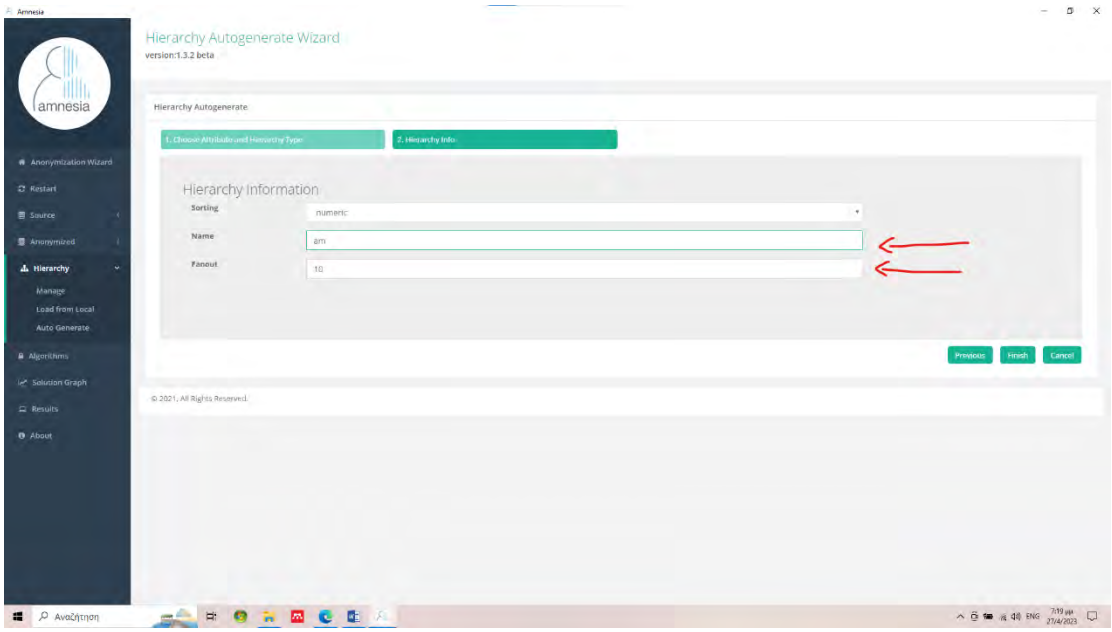


Εικόνα 31: Επιλογή χαρακτηριστικού και τύπου.

Στη συνέχεια επιλέγουμε είδος ταξινόμησης (Sorting) κατά αριθμό, γράμμα ή τυχαία σειρά, όνομα ιεραρχίας και εύρος ομαδοποίησης (Fanout) (βλ. σελ. 38 και εικόνα 12 του παρόντος). Δηλαδή πόσες παρατηρήσεις από τη μεταβλητή AM θα ομαδοποιηθούν ώστε να αποτελέσουν το κάθε επίπεδο γενίκευσης.

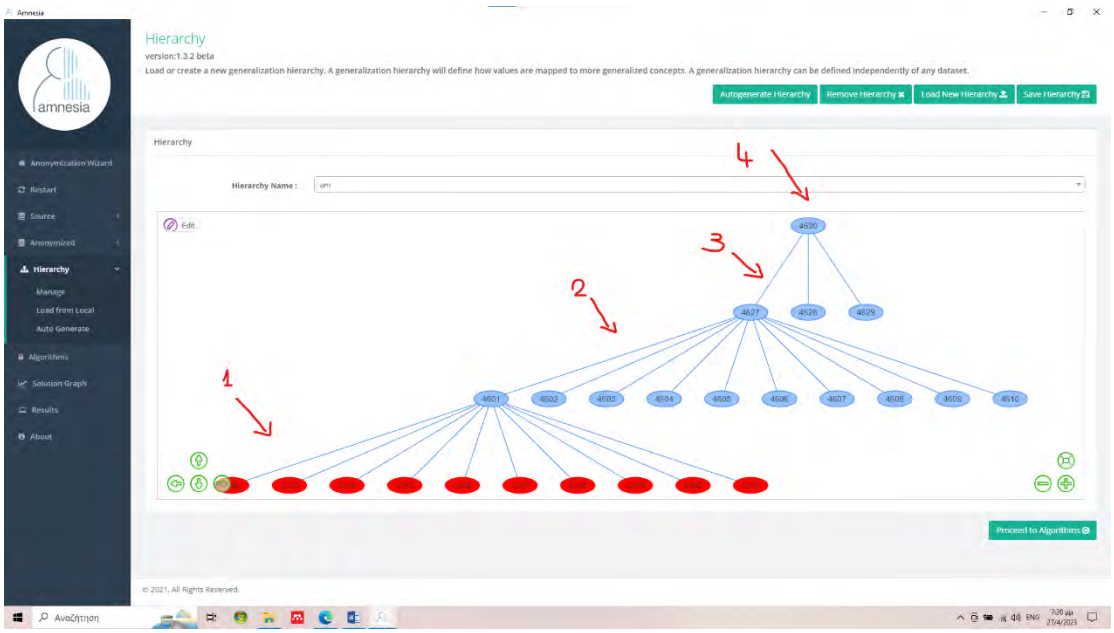


Εικόνα 32: Κανόνες ιεραρχίας. Ταξινόμηση και ομαδοποίηση. Εικόνα α



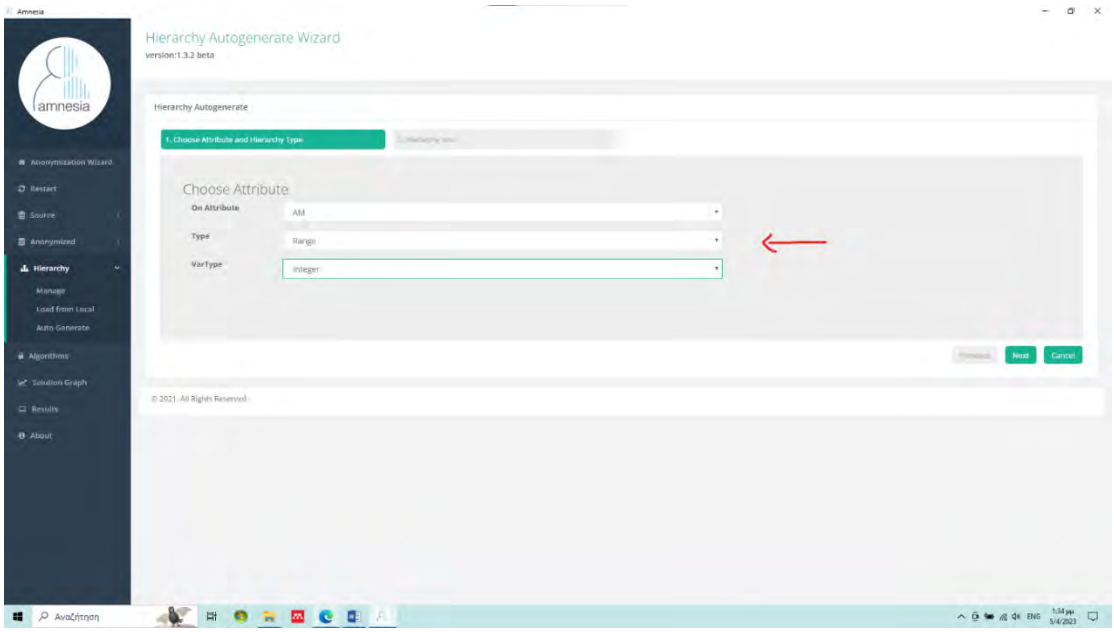
Εικόνα 33: Κανόνες ιεραρχίας. Ταξινόμηση και ομαδοποίηση. Εικόνα β

Παρατηρούμε την εφαρμογή του εύρους ομαδοποίησης (fanout = 10), όπου για κάθε 10 μεταβλητές έχουμε γενίκευση σε κόμβο υψηλότερου επιπέδου (κόκκινα σύννεφα – κλαδιά του δέντρου). Αντίστοιχα στο δεύτερο επίπεδο για κάθε δέκα κόμβους έχουμε ξανά γενίκευση σε κόμβο τρίτου επιπέδου (κορμός του δέντρου). Λογική κατάληξη είναι πως στο τελευταίο επίπεδο (το τέταρτο επίπεδο) δεν μπορούν άλλο να ομαδοποιηθούν οι μεταβλητές και καταλήγουμε στον τελικό κόμβο που αποτελεί τη γενικευμένη μεταβλητή του συγκεκριμένου χαρακτηριστικού (ρίζα του δέντρου) (βλ. εν. 1.5). Επίσης παρατηρούμε ότι κάθε «σύννεφο» χαρακτηρίζεται διακριτά από έναν αριθμό.

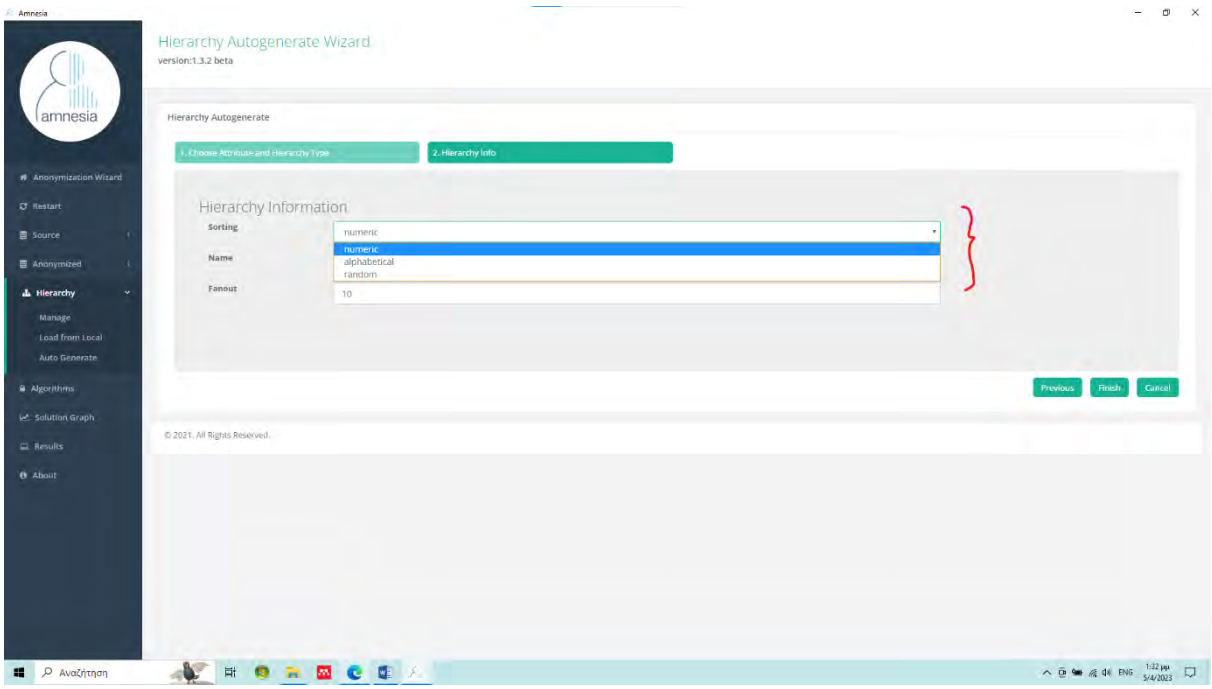


Εικόνα 34: Ομαδοποίηση και γενίκευση του επιλεγέντος χαρακτηριστικού.

- αν θέλουμε να εμφανίζεται το εύρος της ομάδας (Range) και ο μεταβλητός τους τύπος (VarType) να είναι ακέραιος αριθμός καθώς πρόκειται για αριθμούς μητρώου. Επιλέγουμε Next

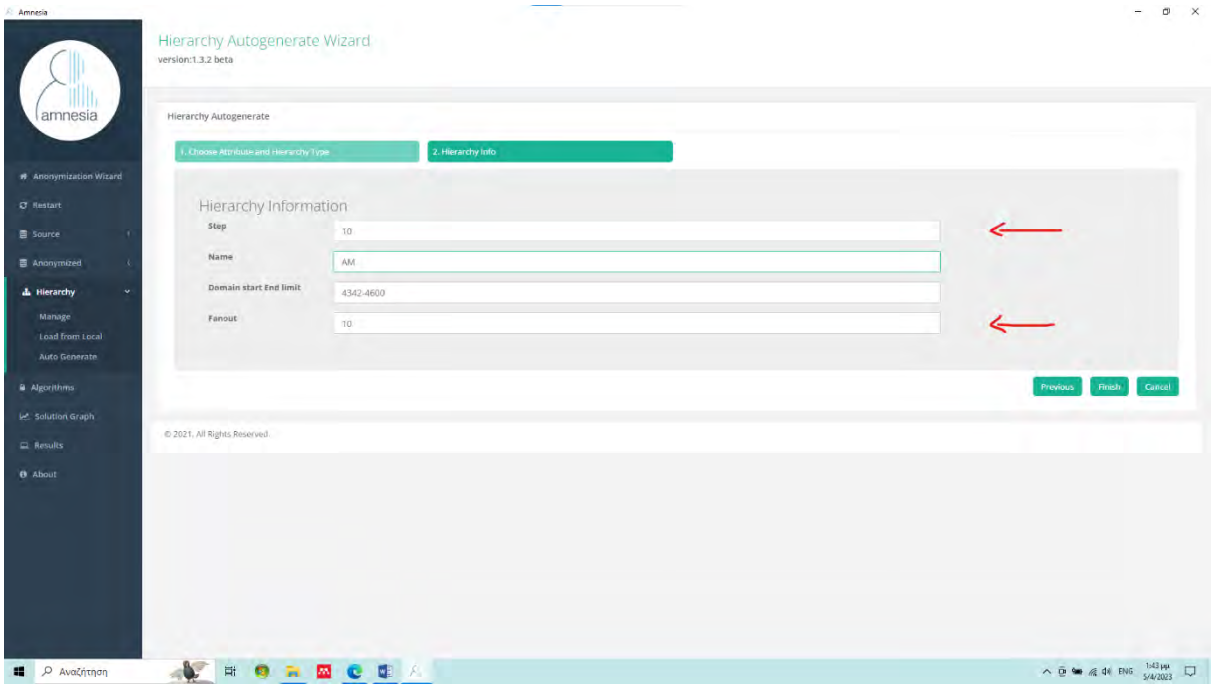


Εικόνα 35: Δημιουργία ιεραρχιών με τις αυτόματες επιλογές του λογισμικού. Εικόνα α

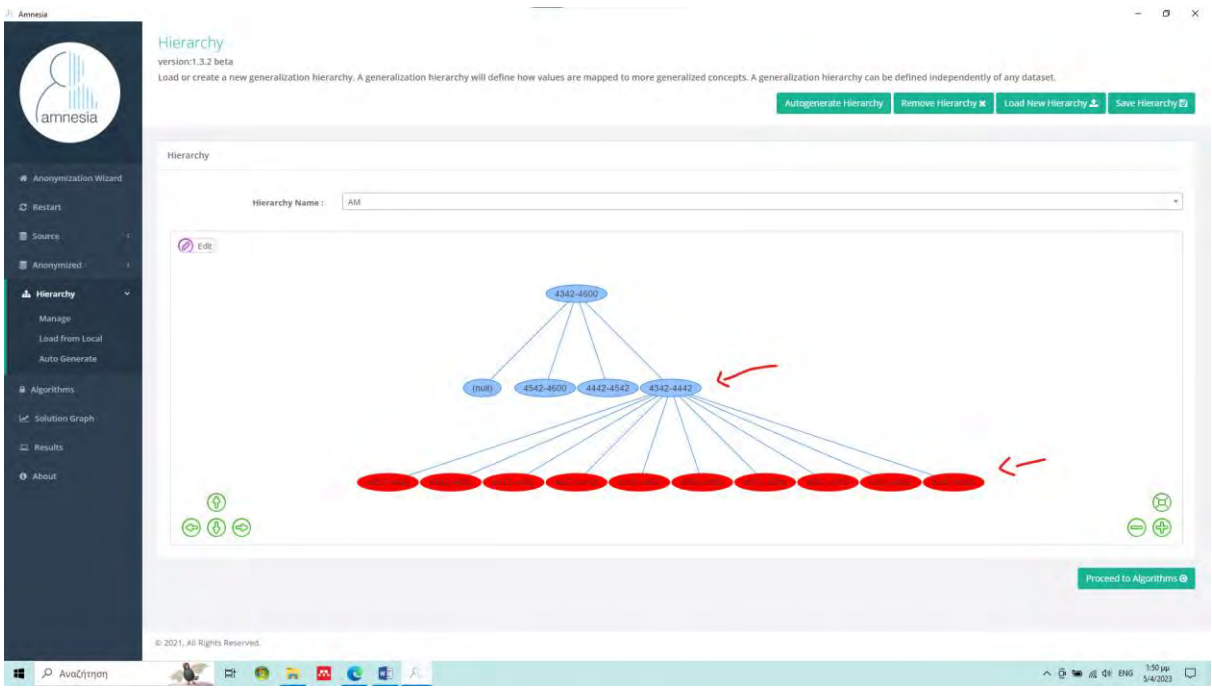


Εικόνα 36: Δημιουργία ιεραρχιών με τις αυτόματες επιλογές του λογισμικού. Εικόνα β

στην επιλογή step αναγράφουμε τον «αριθμό βήματος» που επιθυμούμε. Δηλαδή με αυτήν την επιλογή αναφέρουμε στο λογισμικό ποιο εύρος δεδομένων θα περιλαμβάνει το κάθε σύννεφο. Δηλώνουμε το όνομα και το σύνολο των δεδομένων που θέλουμε να συμπεριλάβουμε ή όποιο σύνολο δεδομένων επιθυμεί ο χρήστης και τέλος το εύρος της ομαδοποίησης (Fanout).



Εικόνα 37: Δημιουργία ιεραρχιών από τις αυτόματες επιλογές του λογισμικού. Εικόνα α



Εικόνα 38: Γενίκευση και ομαδοποίηση χαρακτηριστικού. Εικόνα β

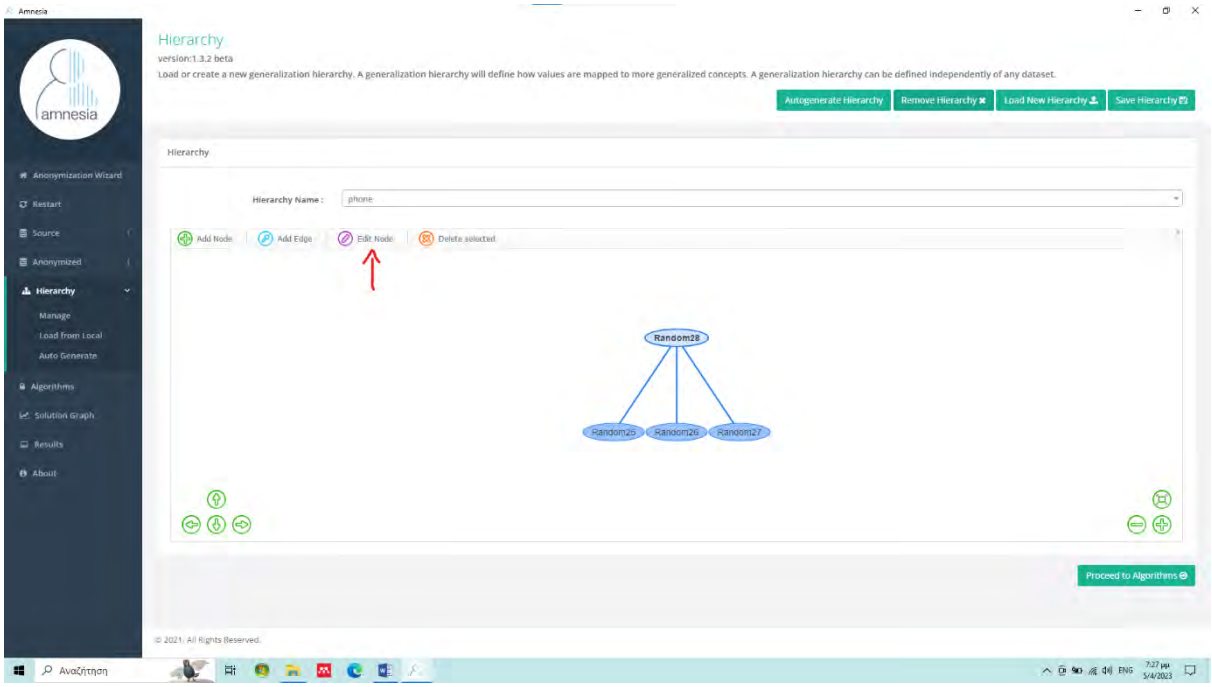
Παρατηρούμε πως εδώ η γενίκευση εφαρμόζεται σε τρία επίπεδα. Υπάρχουν και εδώ από κάτω προς τα πάνω κόκκινα σύννεφα (κλαδιά του δέντρου), το επόμενο επίπεδο μπλε σύννεφα

(κορμός δέντρου) και το τελικό μπλε σύννεφο (ρίζα του δέντρου). Τα κόκκινα σύννεφα έχουν εύρος δέκα βημάτων και το εύρος ομαδοποίησης είναι επίσης δέκα.

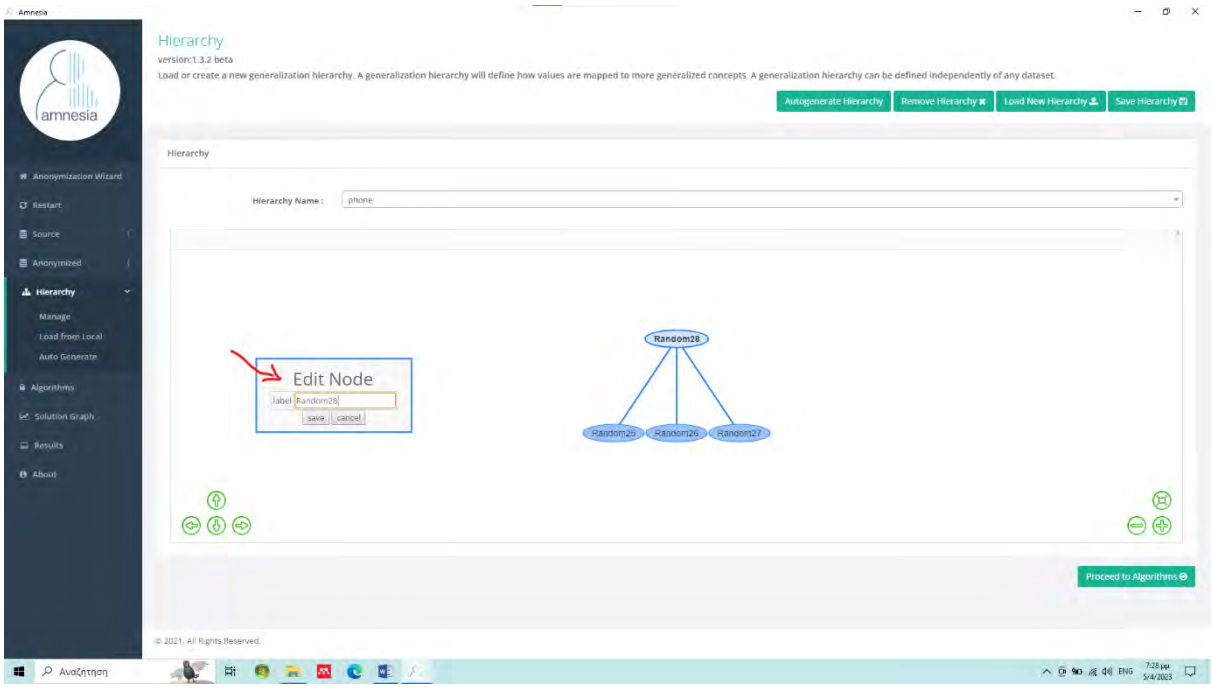
Επιπρόσθετα μπορούμε να αλλάξουμε τους κόμβους γενίκευσης σε οποιοδήποτε όνομα επιθυμούμε με το εικονίδιο edit.

Μας ανοίγουν διάφορες επιλογές με τις οποίες μπορούμε να επέμβουμε με ποικίλους τρόπους πάνω στους κόμβους:

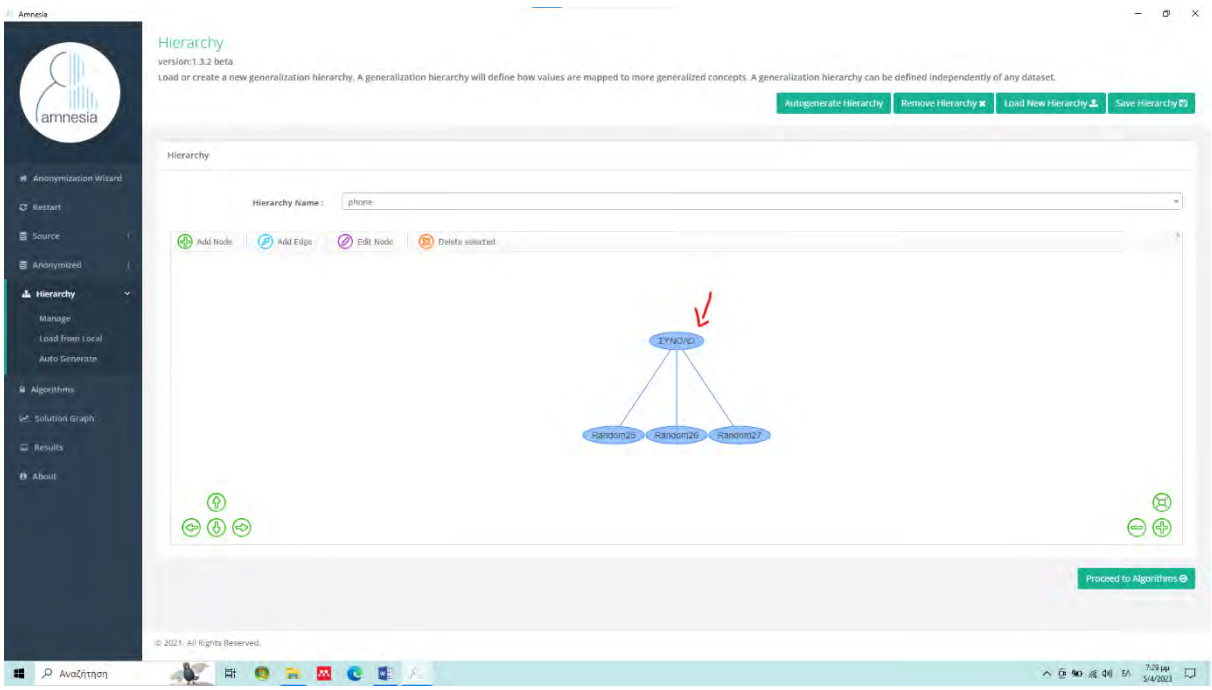
- Add node: μπορούμε να εισάγουμε νέο κόμβο
- Add edge: μπορούμε να ενώσουμε το νέο κόμβο με τον τελικό κόμβο
- Edit node: μπορούμε να μετονομάσουμε τον κάθε κόμβο
- Delete selected: διαγραφή οποιουδήποτε κόμβου



Εικόνα 39: Επεξεργασία κόμβου.



Εικόνα 40: Επεξεργασία-Μετονομασία κόμβου. Εικόνα α



Εικόνα 41: Επεξεργασία-Μετονομασία κόμβου. Εικόνα β

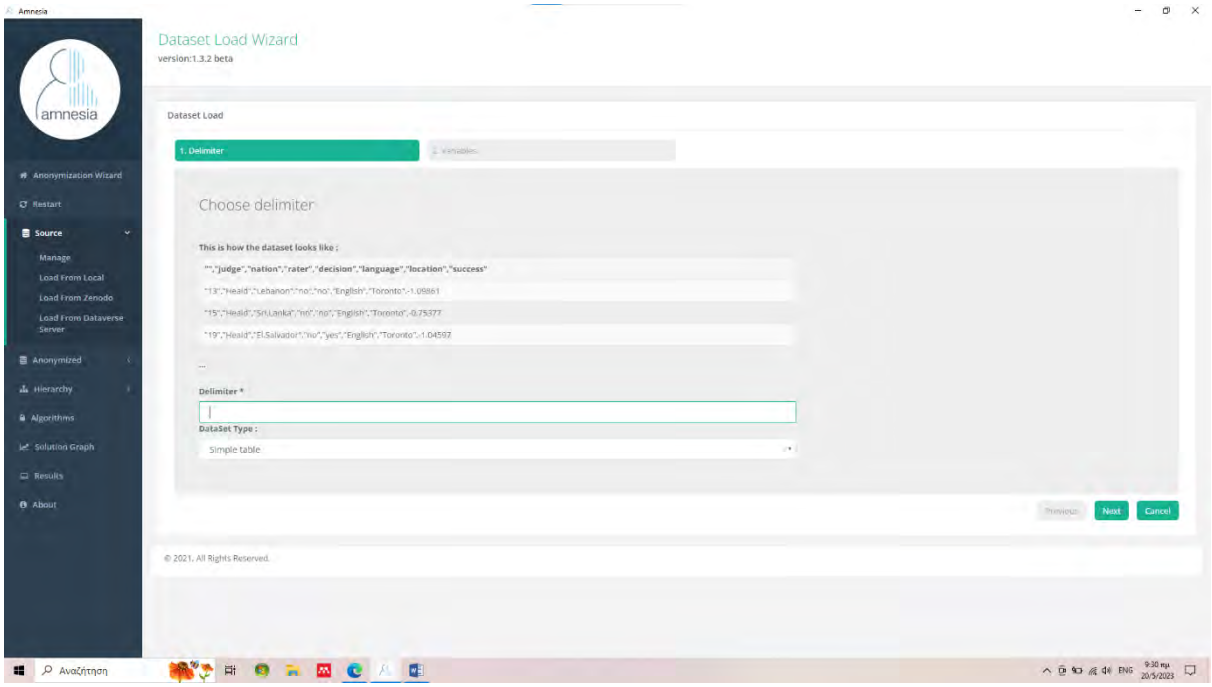
Τέλος, επιλέγουμε την ένδειξη Save Hierarchy για να αποθηκεύσουμε τους κανόνες ιεραρχίας που δημιουργήσαμε, σε αρχείο με κατάληξη .txt, οπουδήποτε στον υπολογιστή μας και να μπορέσουμε να τους φορτώσουμε στο επόμενο βήμα.

Ομοίως πράττουμε και για το χαρακτηριστικό AGM.

3.4.2 Εφαρμογή σε ποιοτικές μεταβλητές

Στην προσπάθειά μας να καταστήσουμε όσο σαφέστερα γίνεται την έννοια των κανόνων ιεράρχησης και το πόσο σημαντικό βήμα είναι αυτό πριν την εφαρμογή οποιωνδήποτε αλγόριθμων, θα παραθέσουμε ένα άλλο παράδειγμα στο οποίο θα δημιουργήσουμε κανόνες ιεραρχίας σε ένα σύνολο δεδομένων που αφορά ενστάσεις προσφύγων στην Αμερική. Θα υποθέσουμε ότι επιθυμούμε να δημοσιεύσουμε τις αρχικές και τελικές αποφάσεις των δικαστών στα πλαίσια του ΓΚΠΔ, ενώ θα γενικεύσουμε τις χώρες προέλευσης των ανθρώπων που έχουν κάνει τις ενστάσεις. Το αρχείο έχει τίτλο Greene.csv και επιλέχθηκε από την ιστοσελίδα

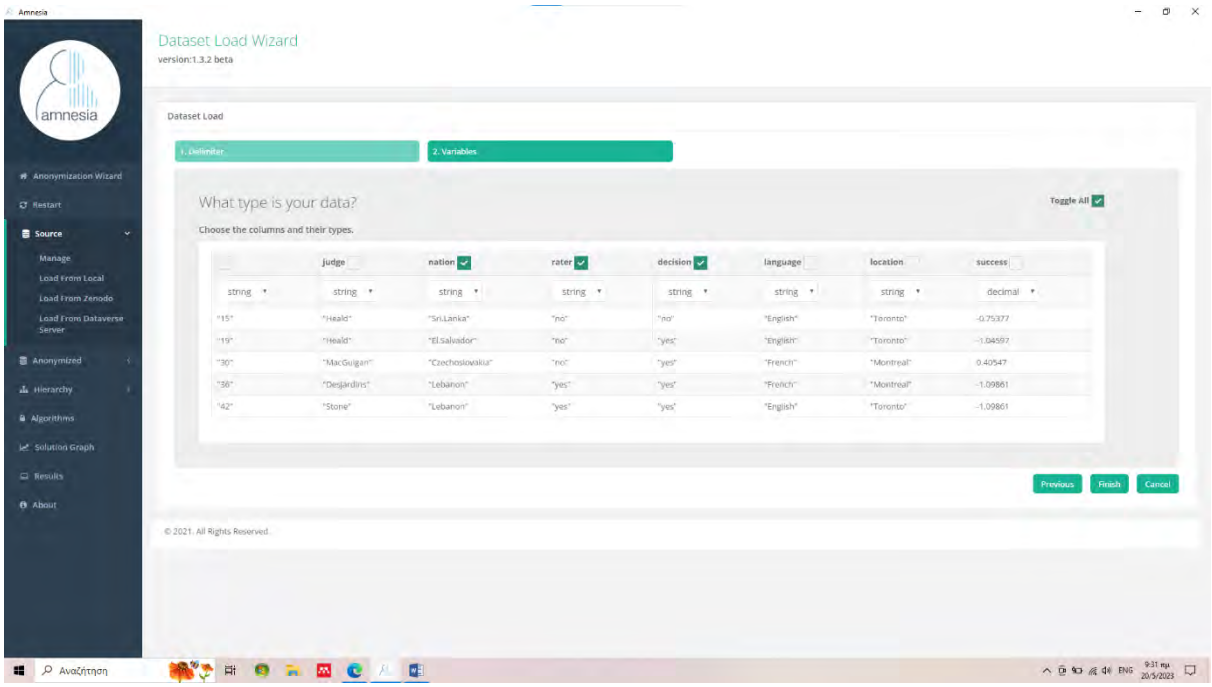
https://vincentarelbundock.github.io/Rdatasets/datasets.html.



Εικόνα 42: Εισαγωγή αρχείου επεξεργασίας Greene.csv

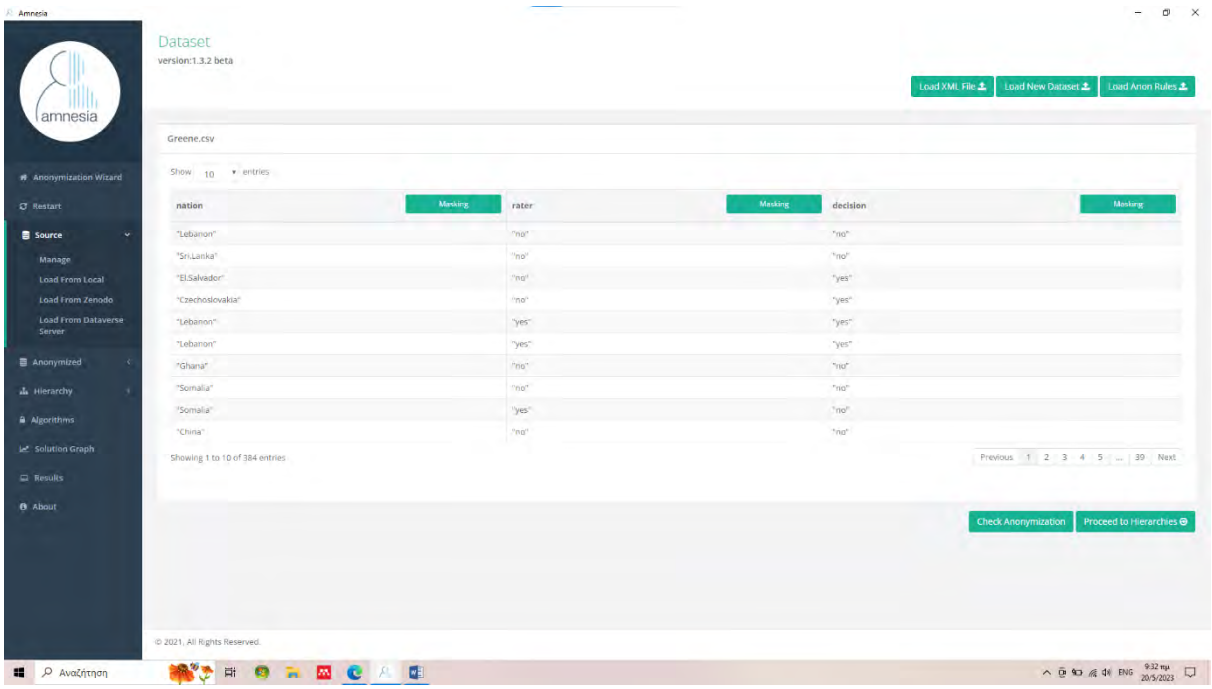
Πρόκειται για ένα σύνολο δεδομένων που απαρτίζεται από τις εξής κατηγορίες:

α) αριθμός φακέλου υπόθεσης, β) όνομα δικαστή, γ) εθνικότητα ενισταμένου, δ) αρχική ετυμηγορία, ε) απόφαση, στ) γλώσσα ενισταμένου, ζ) έδρα εξέτασης, η) ποσοστό επιτυχίας



Εικόνα 43: Επιλογή κατηγοριών που θα εργαστούμε

Για τις ανάγκες αυτής της εργασίας θα επιλέξουμε να ιεραρχήσουμε τις χώρες των ενισταμένων και θα κρατήσουμε τις αποφάσεις ώστε να μπορέσουμε να εξάγουμε αποτελέσματα χωρίς να συνδεθούν με τις χώρες από τις οποίες προέρχονται οι αιτηθέντες την ένσταση.



Εικόνα 44: Επισκόπηση κατηγοριών

Στην συνέχεια θα δημιουργήσουμε ένα αρχείο .txt με τους κανόνες ιεράρχησης που επιθυμούμε να εφαρμόσουμε στο σύνολο δεδομένων Greene.csv.

hierarchies2.txt - Σημειωματάριο

Αρχείο Επεξεργασία Μορφή Προβολή Βοήθεια

```
distinct
name nation
type string
height 3

Africa has Ghana Nicaragua Nigeria Somalia
Asia has China Fiji India Iran Lebanon Pakistan Sri.Lanka
Europe has Bulgaria Czechoslovakia Poland
NorthAmerica has Argentina El.Salvador Guatemala

Continent has Africa Asia Europe NorthAmerica
```

Εικόνα 45: Το αρχείο κανόνων ιεραρχίας που δημιουργήσαμε σε μορφή .txt

Παρατηρούμε ότι έχουμε δώσει τα εξής χαρακτηριστικά:

α) διακριτές τιμές, β) όνομα = nation (εθνικότητα), γ) τύπος = string (συμβολοσειρά) και

δ) ύψος = 3

Στις επόμενες σειρές δηλώνουμε πως επιθυμούμε να ομαδοποιήσουμε τις χώρες ανάλογα με την ήπειρο στην οποία ανήκουν. Επομένως έχουμε την

Α) Αφρική (Γκάνα, Νικαράγουα, Νιγηρία, Σομαλία),

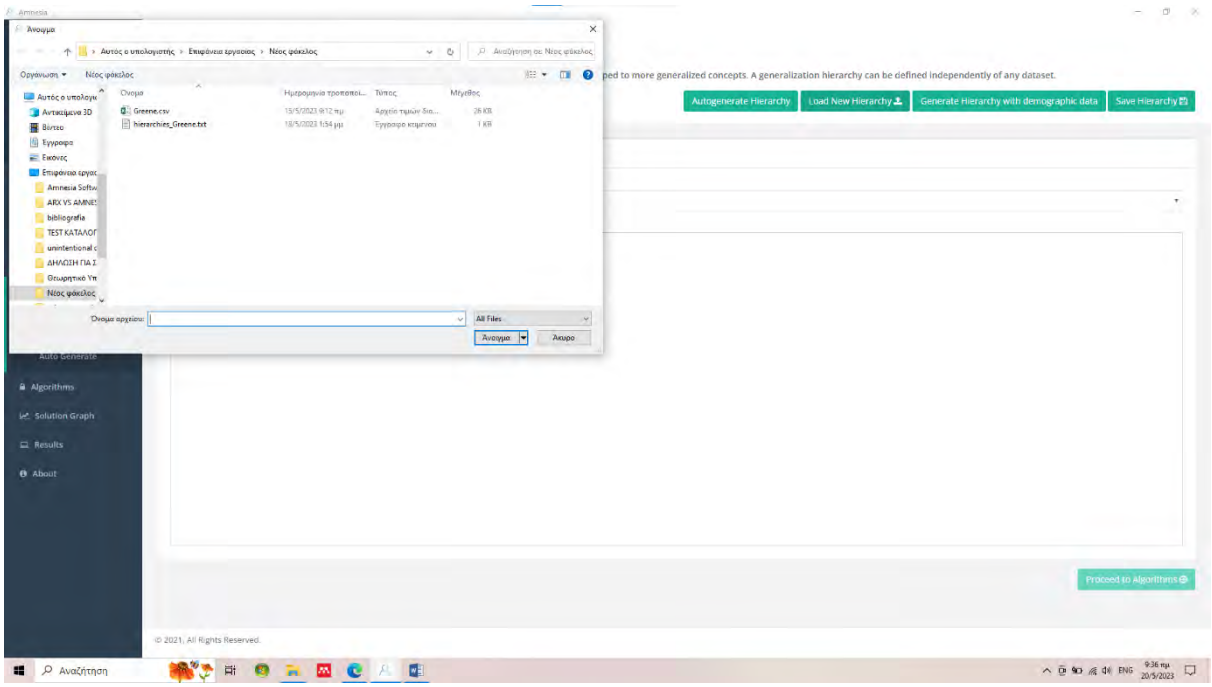
Β) Ασία (Κίνα, Φίτζι, Ινδία, Ιράν, Λίβανος, Πακιστάν, Σρι Λάνκα),

Γ) Ευρώπη (Βουλγαρία, Τσεχοσλοβακία, Πολωνία) και

Δ) Αμερική (Αργεντινή, Ελ Σαλβαδόρ, Γουατεμάλα).

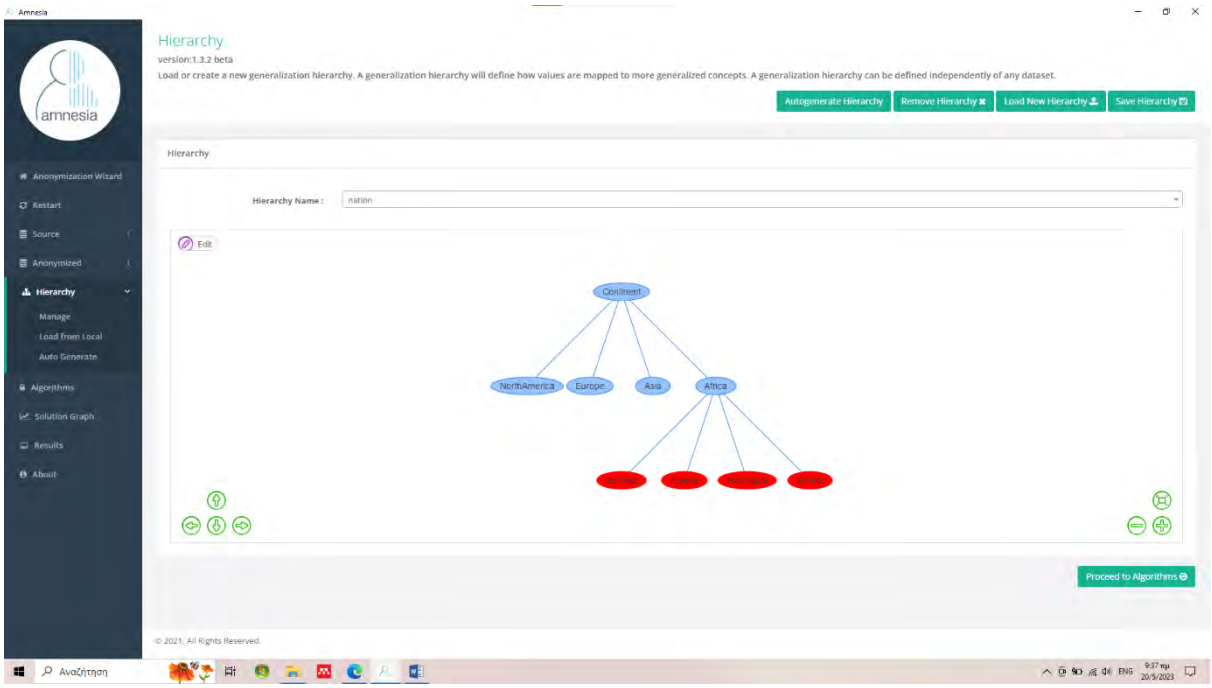
Τελευταίο τρίτο επίπεδο έχουμε την τελική ομαδοποίηση όλων των χωρών και των ηπείρων ως «Continent».

Φορτώνουμε τους κανόνες ιεράρχησης από το αντίστοιχο μενού του λογισμικού:

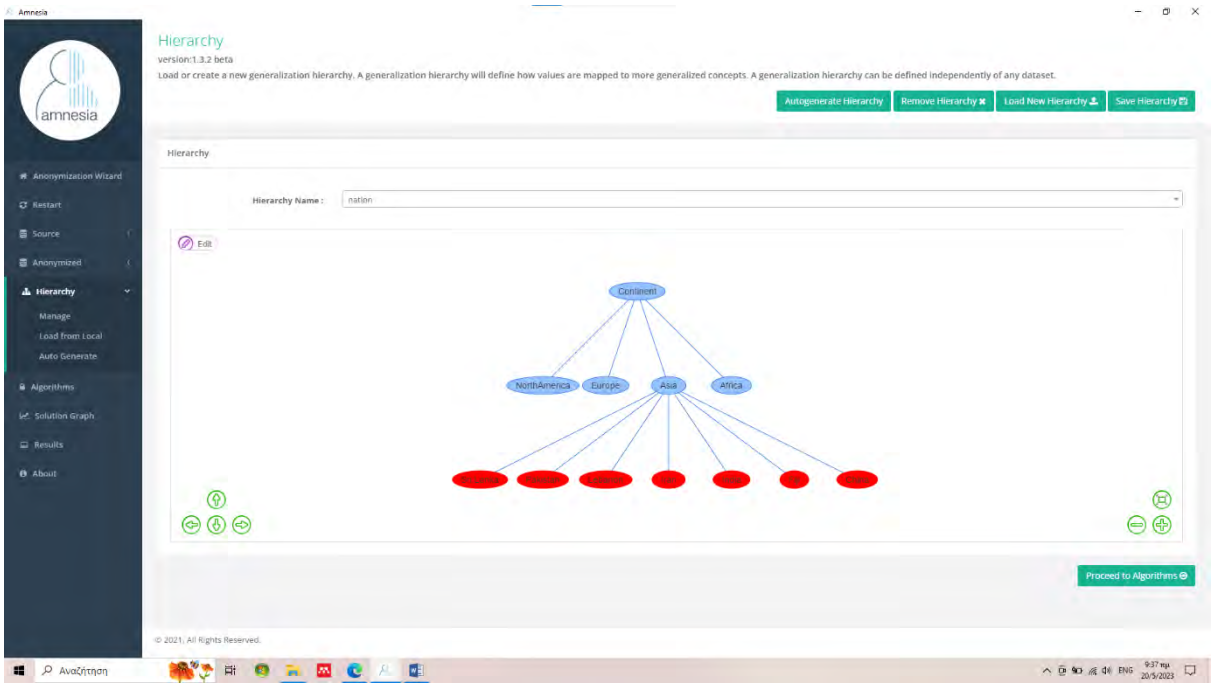


Εικόνα 46: Εισαγωγή αρχείου ιεράρχησης με μορφή .txt

και παρατηρούμε πως έχουν εφαρμοστεί οι κανόνες ιεράρχησης που έχουμε δημιουργήσει:



Εικόνα 47: Απεικόνιση ιεράρχησης σύμφωνα με τους κανόνες που δημιουργήσαμε για τις χώρες της Αφρικής

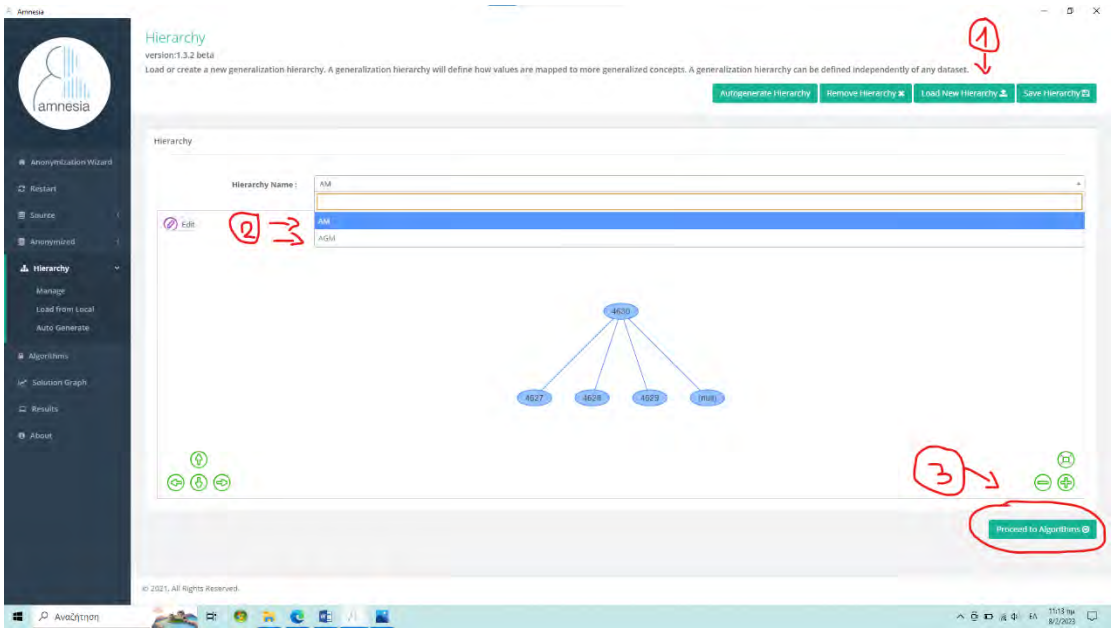


Εικόνα 48: Απεικόνιση ιεράρχησης σύμφωνα με τους κανόνες που δημιουργήσαμε για τις χώρες της Ασίας

Το πρώτο επίπεδο αποτελείται από τις αρχικές ομάδες, οι οποίες ομαδοποιούνται ανά ήπειρο στο δεύτερο επίπεδο, ενώ στο τρίτο τελικό επίπεδο (άρα επαληθεύεται και το height = 3), έχουμε την ομαδοποίηση με κανόνες ιεράρχησης στο μπλε σύννεφο «Continent» όλων ανεξαιρέτως των δηλωθέντων χωρών.

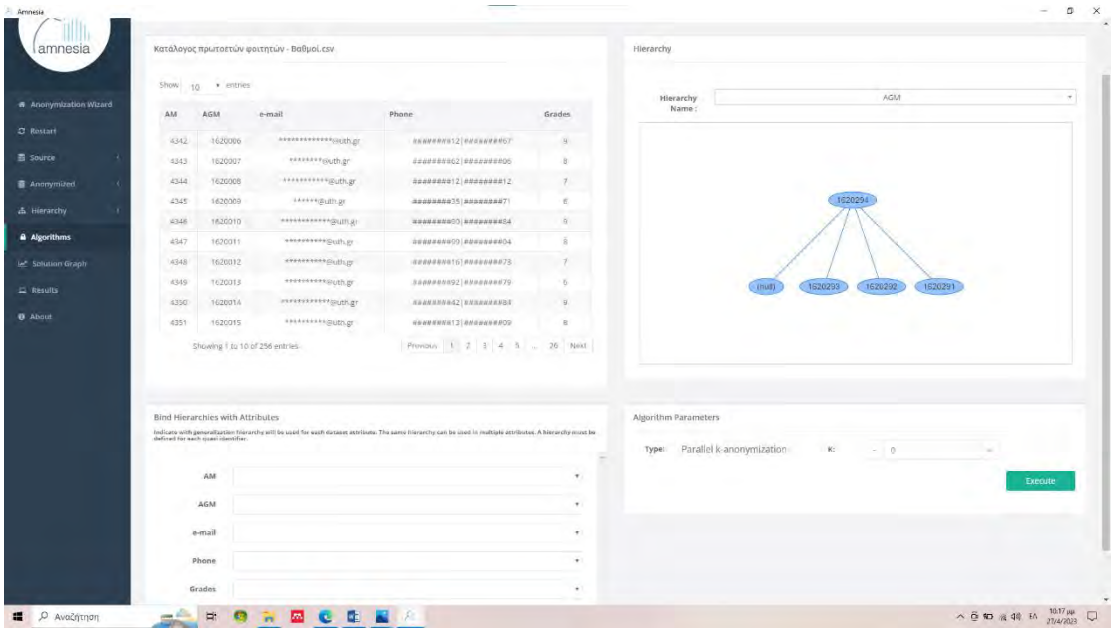
3.5 Εφαρμογή Αλγόριθμων ανωνυμοποίησης

Φορτώνουμε μέσω της επιλογής Load New Hierarchy, από τη θέση που είχαμε προηγούμενα αποθηκεύσει, τα δύο αρχεία με κατάληξη .txt (με τις μεταβλητές AM και AGM) που περιέχουν τους κανόνες ιεραρχίας που δημιουργήσαμε και προχωράμε στην επιλογή Proceed to Algorithms.

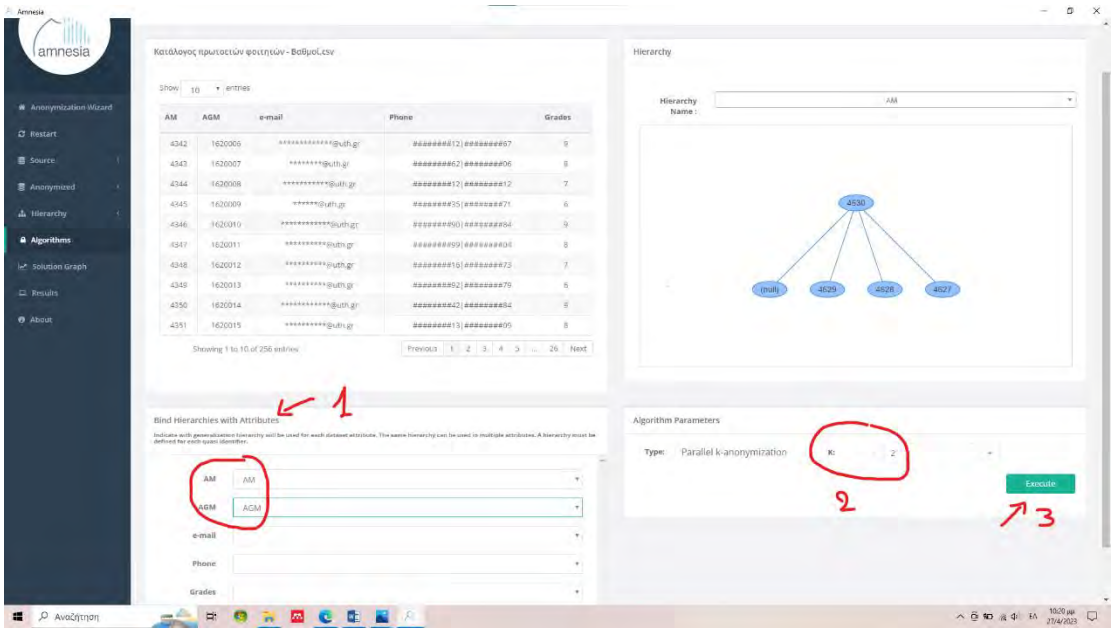


Εικόνα 49:Εισαγωγή αλγόριθμων ανωνυμοποίησης.

Παρατηρούμε ότι έχουμε τις στήλες όλων των χαρακτηριστικών με εμφανή την συγκάλυψη στα email και phone. Στο επόμενο βήμα θα πρέπει με τη χρήση του αλγόριθμου k-ανωνυμίας να «δέσουμε» ιεραρχίες (Bind Hierarchies) με τις εναπομείνασες μεταβλητές AM και AGM, επιλέγοντας για το κάθε χαρακτηριστικό την αντίστοιχη ιεραρχία που έχουμε δημιουργήσει. Επίσης επιλέγουμε k=2 γιατί επιθυμούμε να υπάρξει ανωνυμοποίηση στις εγγραφές με τουλάχιστον 2 ίδια οιονεί χαρακτηριστικά, στο σύνολο των γενικευμένων δεδομένων μας. Επιλέγουμε Execute.



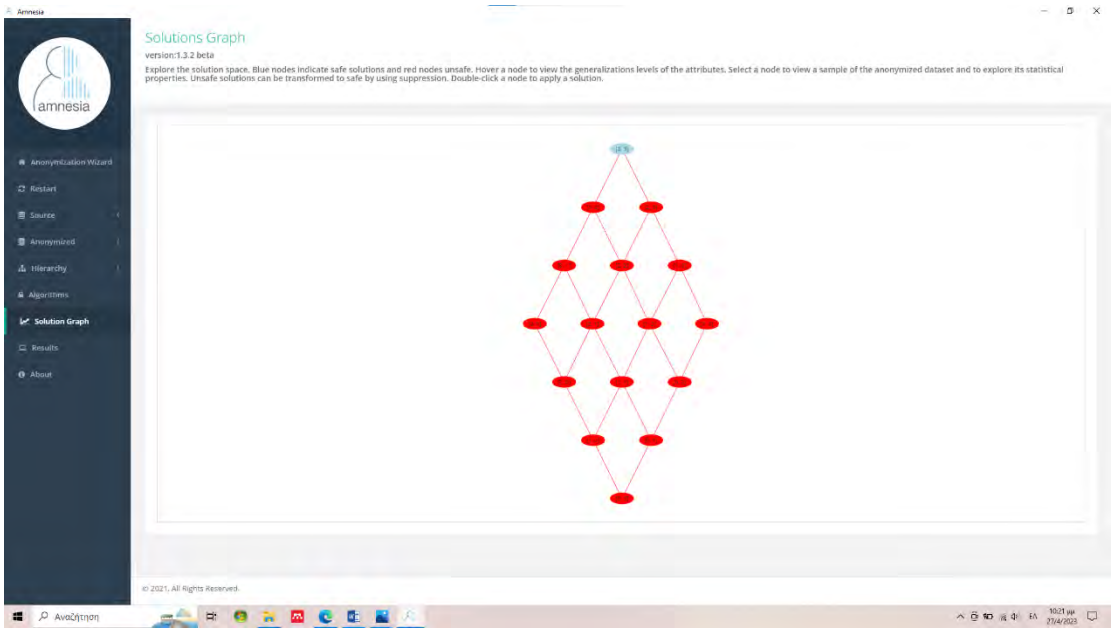
Εικόνα 50: Το "δέσιμο" των αλγόριθμων και των χαρακτηριστικών. εικόνα α



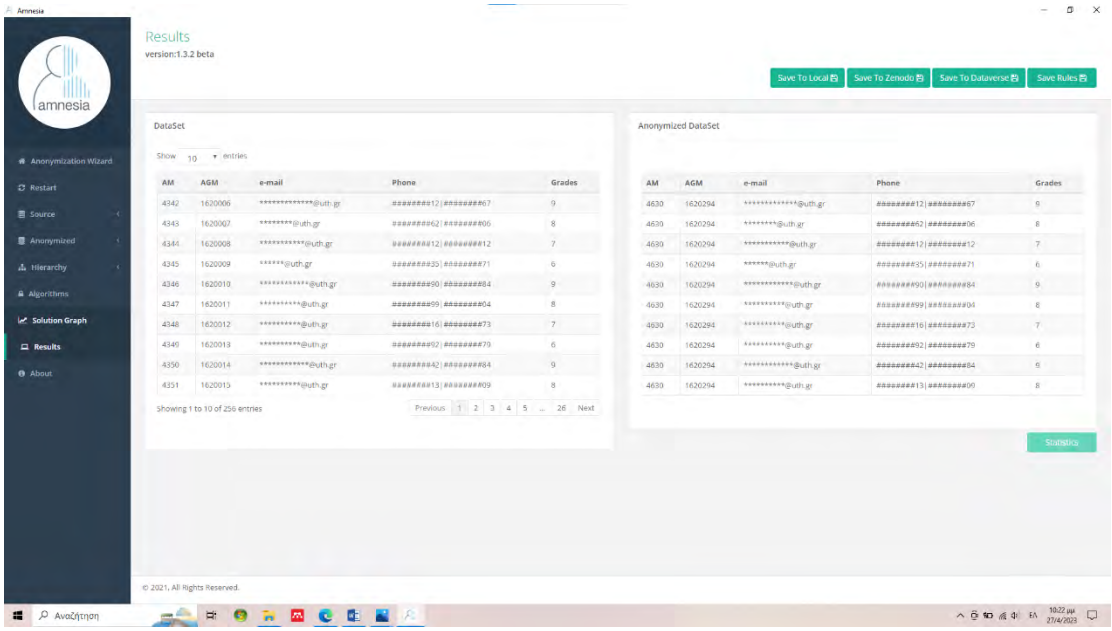
Εικόνα 51: Το "δέσιμο" των αλγόριθμων και των χαρακτηριστικών. Εικόνα β

3.6 Γράφημα λύσης

Σε αυτό το τελικό στάδιο αφού συνδέσαμε μέσω του αλγόριθμου k-ανωνυμίας τις ιεραρχίες με τα χαρακτηριστικά, παρατηρούμε πάλι μπλε και κόκκινα σύννεφα που ενώνονται μεταξύ τους. Θα τους ονομάσουμε κόμβους και αποτελούν συνένωση μεταβλητών και ιεραρχιών του αρχικού συνόλου δεδομένων μας. Οι κόκκινοι κόμβοι μας υποδεικνύουν μη ασφαλή ανωνυμοποιημένο συνδυασμό γιατί τα δεδομένα έχουν και άλλα επίπεδα γενίκευσης μέχρι να φτάσουμε στην «ρίζα του δέντρου». Οι μπλε κόμβοι μας υποδεικνύουν πως όλες οι γενικεύσεις έχουν ολοκληρωθεί με τους κανόνες ιεραρχίας που δημιουργήσαμε και προσφέρουν εγγυημένη ασφάλεια. Σταματώντας τον κέρσορά μας πάνω σε κάθε κόμβο, το λογισμικό είναι σε θέση να μας ενημερώσει το επίπεδο γενίκευσης που έχει επιτευχθεί. Κάνοντας διπλό κλικ σε κάθε κόμβο μπορούμε να μελετήσουμε τα σημεία που δεν εγγυόνται την ασφάλεια ανωνυμίας, ενώ επιλέγοντας τον τελικό κόμβο, μας ανοίγει ένα διπλό γραφικό περιβάλλον όπου στο αριστερό μέρος φαίνεται το σύνολο δεδομένων πριν την ανωνυμοποίηση ενώ στο δεξί μέρος το πλήρως ανωνυμοποιημένο και έτοιμο για οποιαδήποτε χρήση σύνολο δεδομένων.



Εικόνα 52: Γράφημα λύσης. Ο μπλε κόμβος είναι η ανωνυμοποιημένη λύση μας.



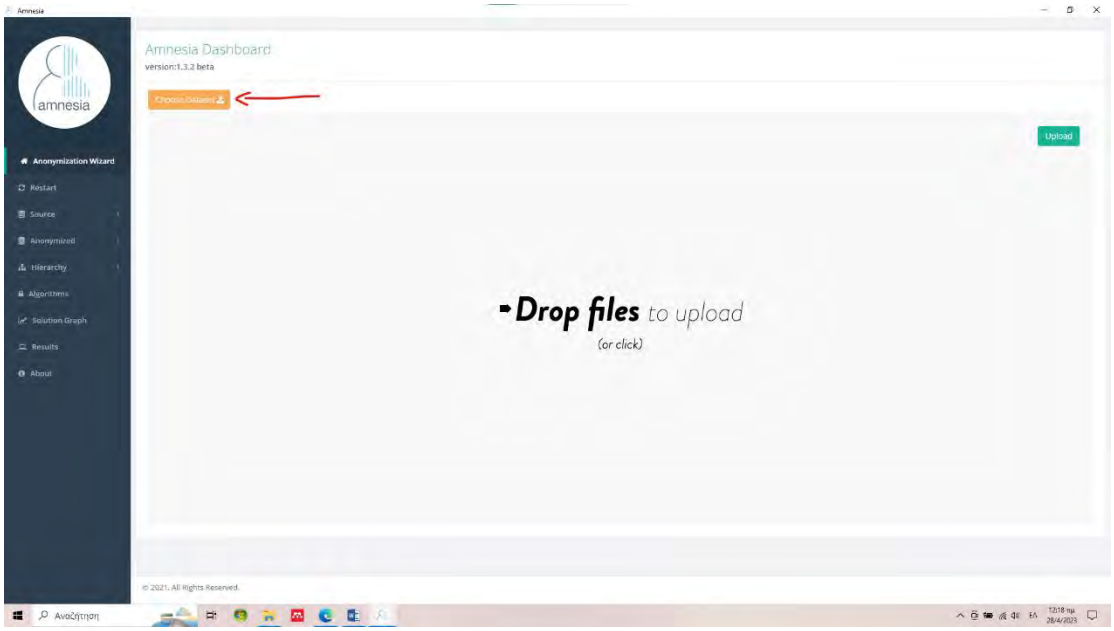
Εικόνα 53: Τελικό αποτέλεσμα. Το αριστερό τμήμα πριν την ανωνυμοποίηση και το δεξί τμήμα πλήρως ανωνυμοποιημένο.

3.7 Εφαρμογή ανωνυμοποίησης με αρχείο .txt

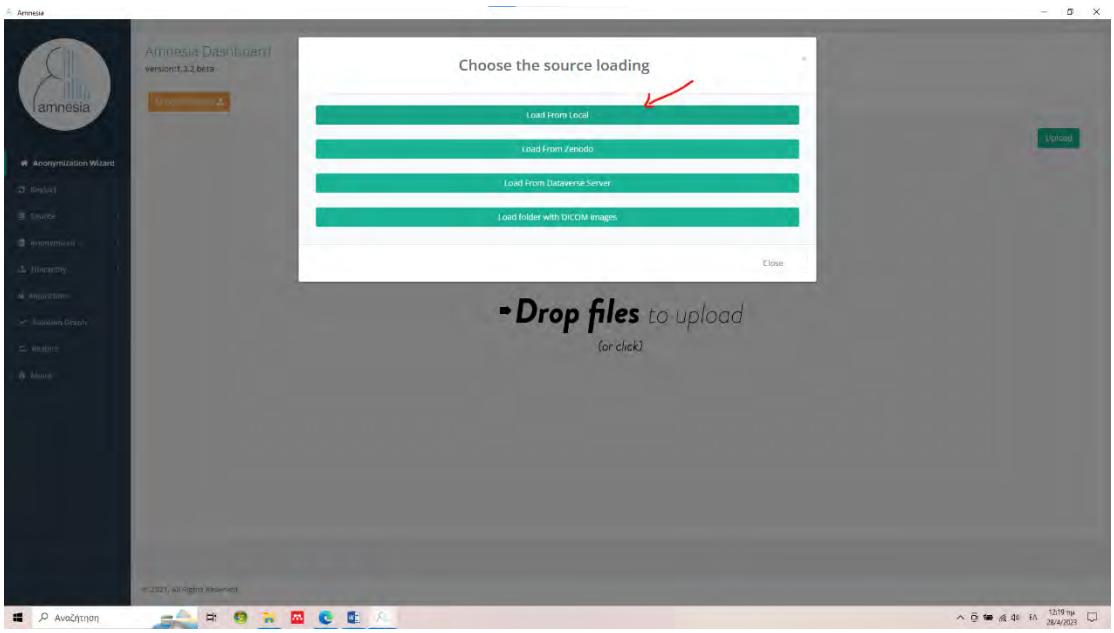
Για τις ανάγκες αυτής της έρευνας θα χρησιμοποιήσουμε ένα σύνολο δεδομένων κατάληξης .txt που αφορά οικονομικά στοιχεία και παρέχεται ως δοκιμαστικό σύνολο δεδομένων (test dataset). Το σύνολο δεδομένων περιλαμβάνει τα εξής: ταχυδρομικό κώδικα (zip), ηλικία (age), αριθμός πιστωτικής κάρτας (credit card), φύλο (gender) και μισθό (salary).

Ο σκοπός μας είναι να μπορέσουμε να ανωνυμοποιήσουμε το σύνολο δεδομένων αφήνοντας εμφανή μόνο το φύλο και το μισθό του κάθε ατόμου.

Αρχεία συνόλου δεδομένων μπορούμε να φορτώσουμε στο λογισμικό είτε στη δοκιμαστική έκδοση (online demo) είτε στον τοπικό μας δίσκο. Επιλέγουμε τη φόρτωση από τον τοπικό μας δίσκο.

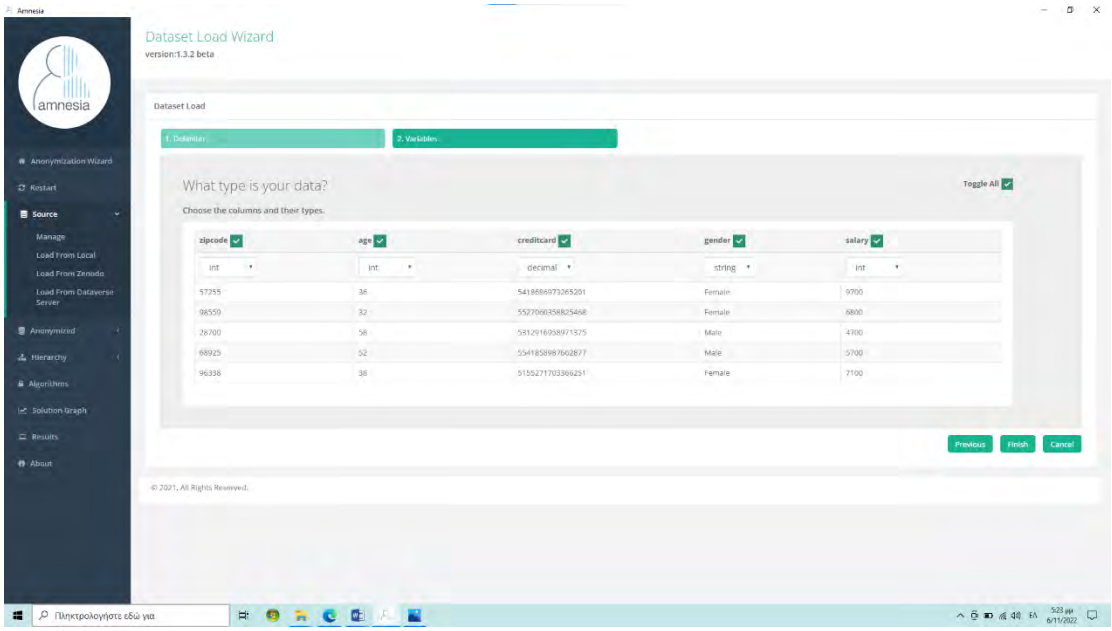


Εικόνα 54:Εισαγωγή αρχείου συνόλου δεδομένων. Εικόνα α

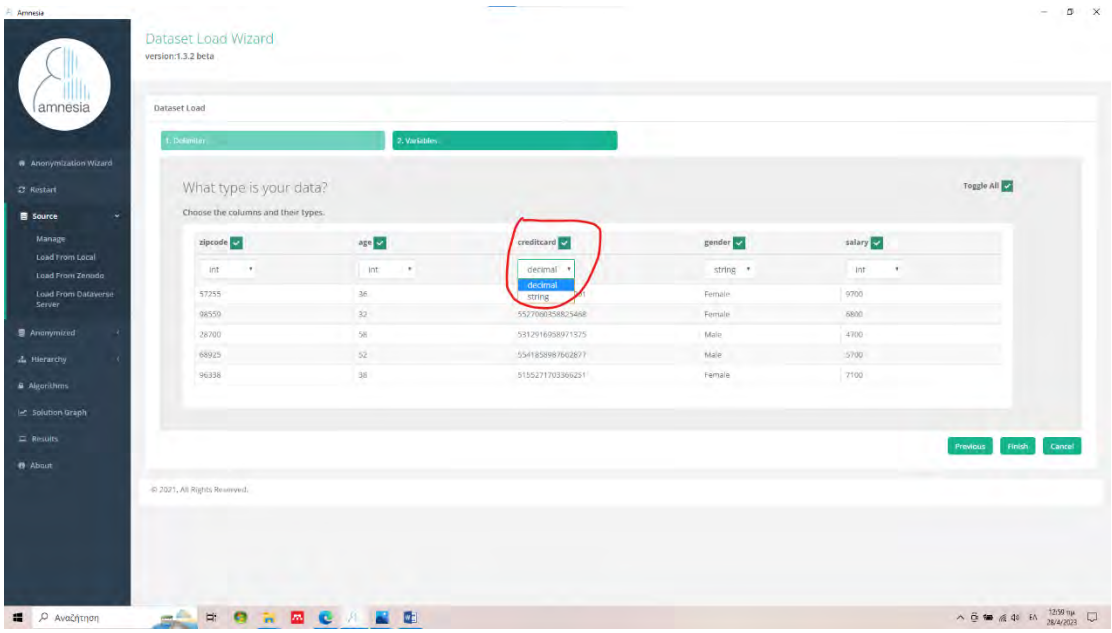


Εικόνα 55: Εισαγωγή αρχείου συνόλου δεδομένων. Εικόνα β

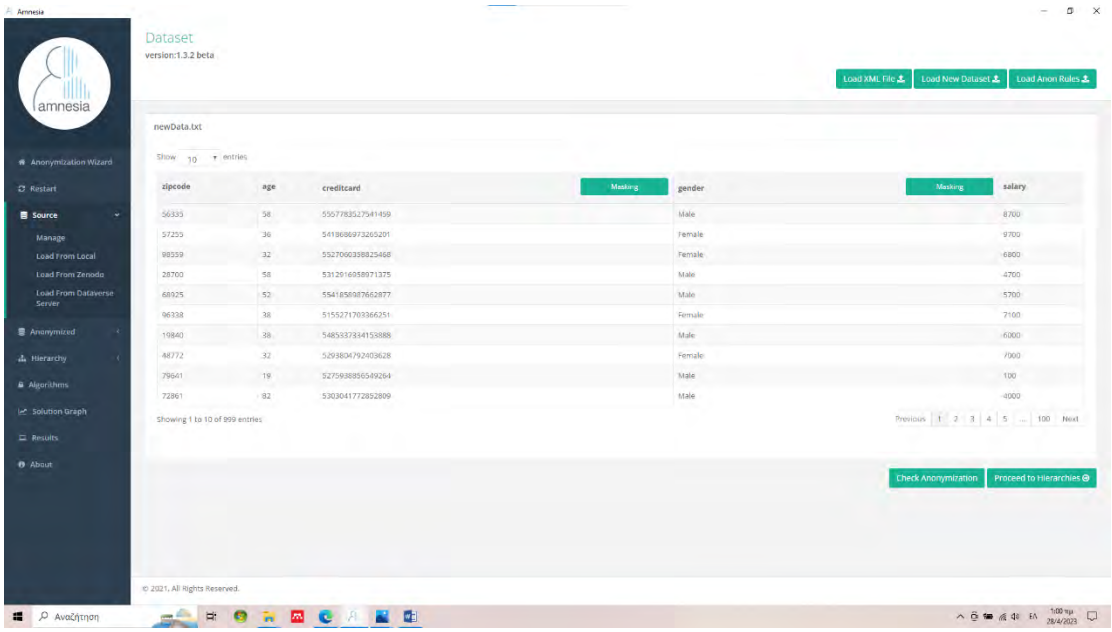
Το λογισμικό είναι σε θέση να αντιληφθεί το είδος των πληροφοριών που περιέχονται, διαβάζοντας τις πρώτες γραμμές του συνόλου δεδομένων που εισάγουμε, ενώ οι όποιες αλλαγές από το χρήστη, γίνονται μόνον εφόσον είναι απαραίτητο (βλ. σελ. 36 του παρόντος). Έτσι στο συγκεκριμένο παράδειγμα, επειδή αντιλαμβάνεται τη μεταβλητή credit card ως δεκαδικό (decimal), θα το αλλάξουμε σε χαρακτήρα σειράς (string).



Εικόνα 56: Αλλαγή τύπου χαρακτηριστικού credit card. Εικόνα α



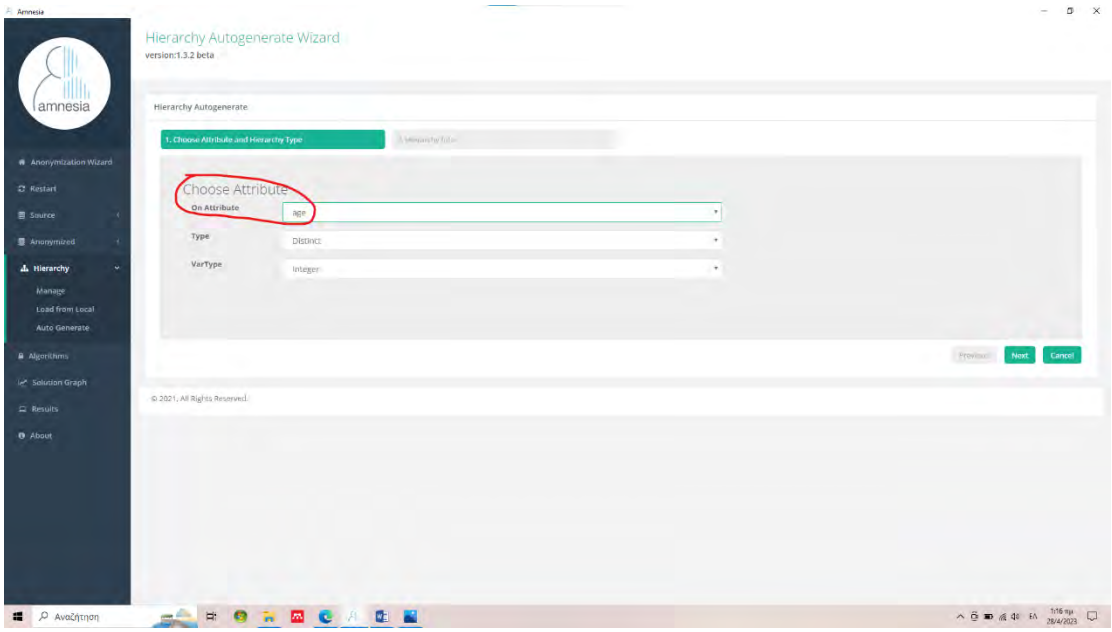
Εικόνα 57: Αλλαγή τύπου χαρακτηριστικού credit card. Εικόνα β



Εικόνα 58: Το σύνολο δεδομένων πριν την δημιουργία ιεραρχίας.

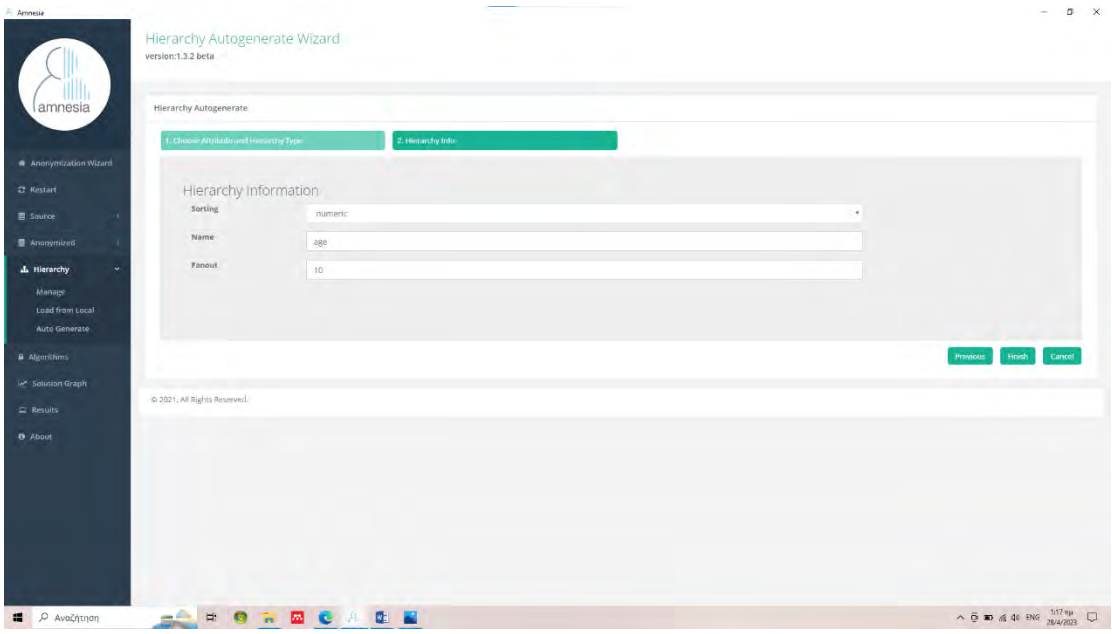
Επειδή θέλουμε για στατιστικούς λόγους να ερευνήσουμε μόνο το φύλο των υπαλλήλων και το εισόδημα που λαμβάνουν θα προχωρήσουμε μέσω των κανόνων ιεραρχίας να γενικεύσουμε τον ταχυδρομικό κώδικα, την πιστωτική κάρτα και την ηλικία.

Επιλέγουμε Proceed to Hierarchies →Autogenerate Hierarchy επιλέγοντας το χαρακτηριστικό ηλικία (age), τύπος διακριτή (Distinct) και μεταβλητός τύπος (VarType) ακέραιος (Integer).

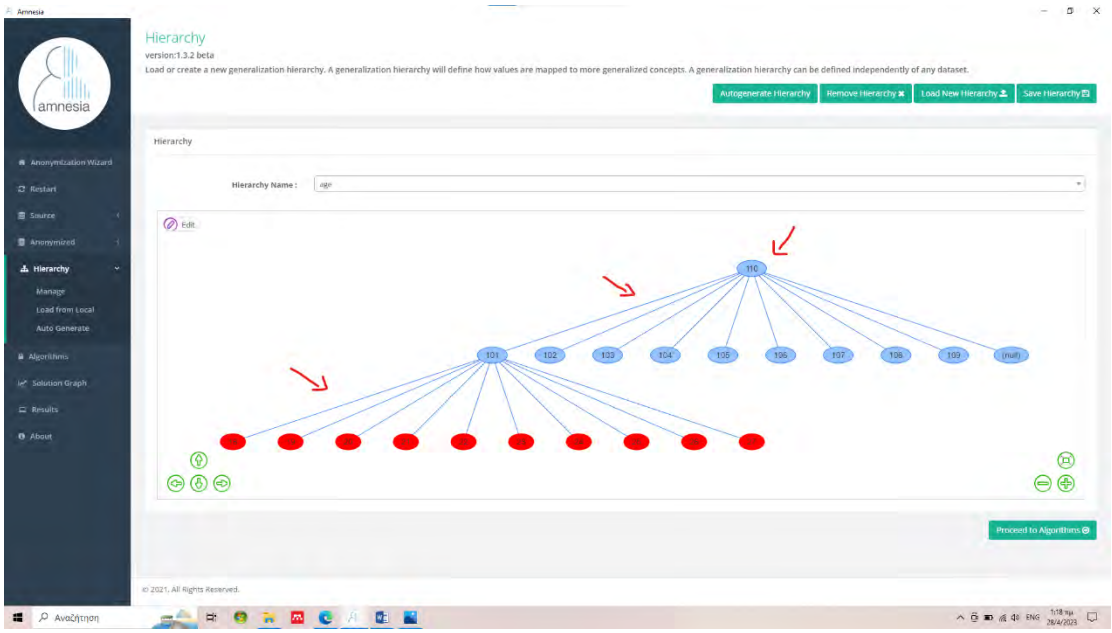


Εικόνα 59: Επιλογή χαρακτηριστικού ηλικίας (age).

Στη συνέχεια επιλέγουμε είδος ταξινόμησης (Sorting) σύμφωνα με το πως θέλουμε να ταξινομηθούν, όνομα κανόνα ιεραρχίας (Name) και εύρος ομαδοποίησης (Fanout).



Εικόνα 60: Απεικόνιση γενίκευσης. Εικόνα α



Εικόνα 61: Απεικόνιση γενίκευσης. Εικόνα β

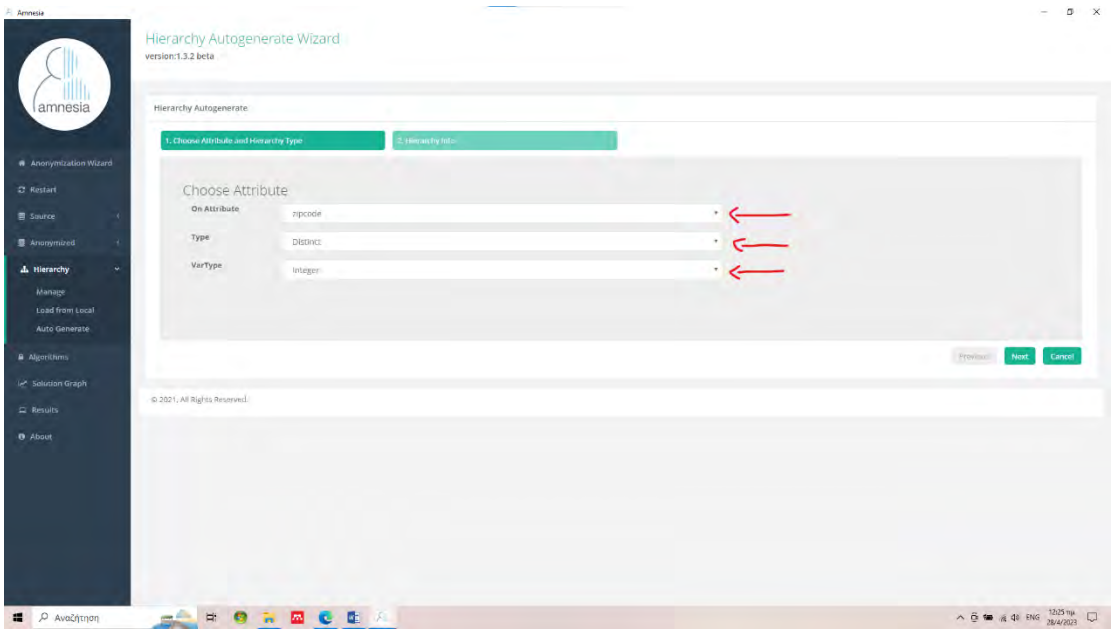
Παρατηρούμε πως το Amnesia έχει επιλέξει να γενικεύσει (ομαδοποιήσει) τις τιμές για ηλικίες από 18 έως 27 στην ομάδα 101, τις τιμές για ηλικίες 28 έως 37 στην ομάδα 102 κοκ. Για να γίνει αντιληπτή η όλη ιδέα της γενίκευσης, αρκεί να αναγάγουμε κάθε αρχική ομάδα

δεδομένων σε μία ανώτερη πιο γενική ομάδα που συμπεριλαμβάνει τα αρχικά δεδομένα. Η πλέον γενικευμένη τιμή (άρα και ανωνυμοποιημένη) είναι η ρίζα του δέντρου, οι πρώτες ομάδες ο κορμός και οι κόκκινες αρχικές τιμές τα κλαδιά.

ΟΜΑΔΑ ΑΡΙΘΜΩΝ (ΚΛΑΔΙΑ)	Α' ΕΠΙΠΕΔΟ ΓΕΝΙΚΕΥΣΗΣ (ΚΟΡΜΟΣ)	Β' ΕΠΙΠΕΔΟ ΓΕΝΙΚΕΥΣΗΣ (ΡΙΖΑ)
8-17	101	110
18-27	102	110
28-37	103	110
38-47	104	110
48-57	105	110
58-67	106	110
68-77	107	110
78-87	108	110
88-97	109	110

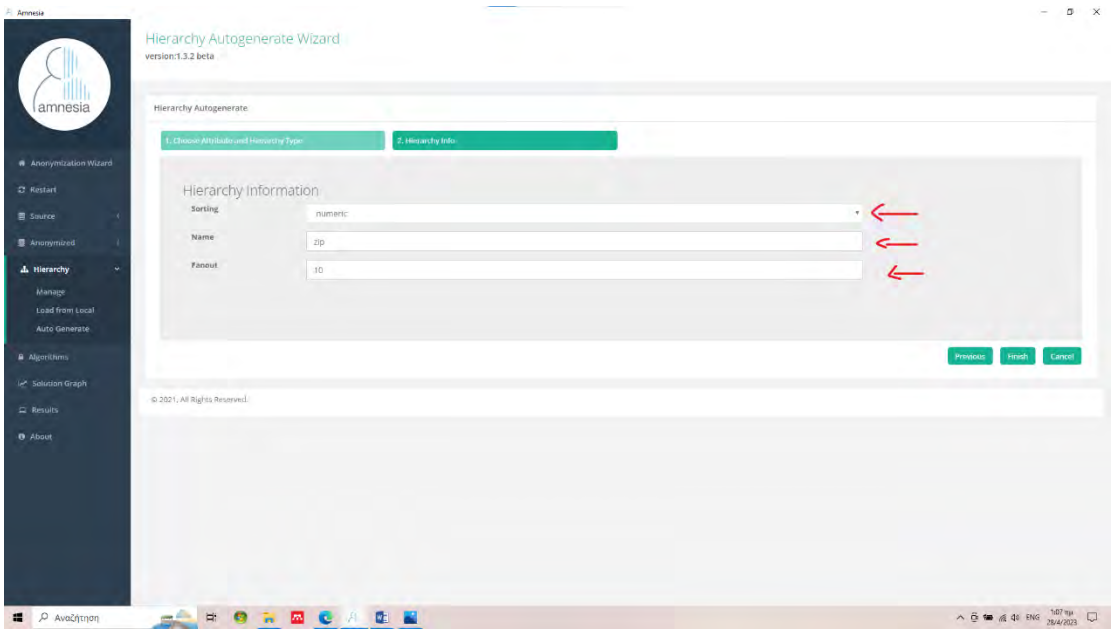
Εικόνα 62: Απεικόνιση γενίκευσης χαρακτηριστικού ηλικία. Ιδίας δημιουργίας.

Για να δημιουργήσουμε κανόνα ιεραρχίας για το χαρακτηριστικό ταχυδρομικός κώδικας (zip code), επιλέγουμε ξανά Autogenerate Hierarchy και εφαρμόζουμε τα ίδια βήματα .



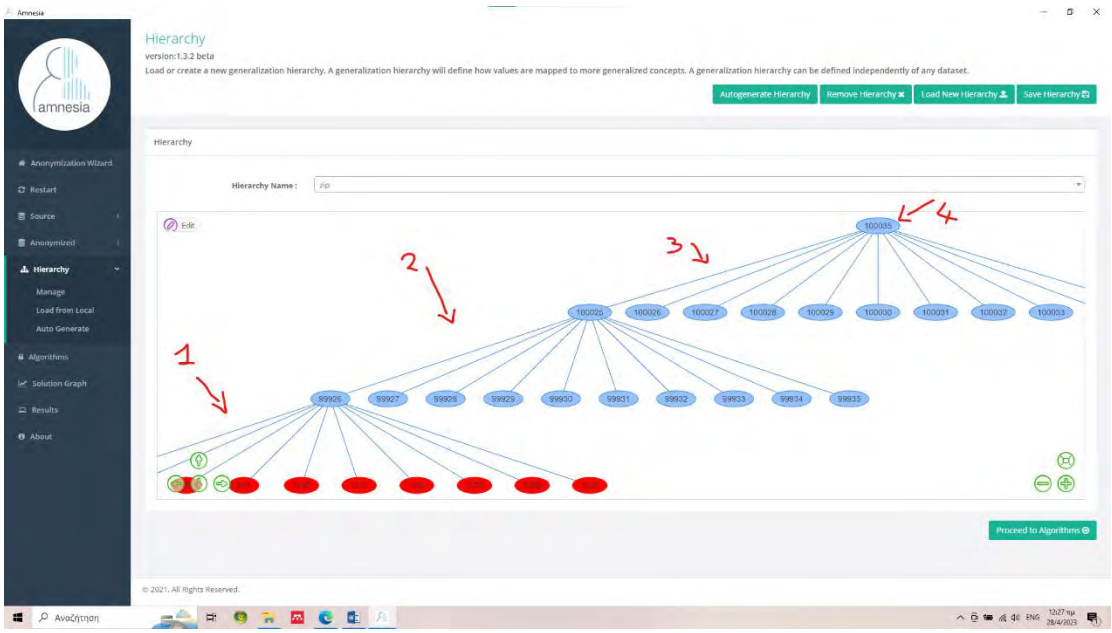
Εικόνα 63: Επιλογή χαρακτηριστικού ταχυδρομικός κώδικας (zipcode)

Στη συνέχεια επιλέγουμε είδος ταξινόμησης (Sorting), όνομα κανόνα ιεραρχίας (Name) και εύρος ομαδοποίησης (Fanout).



Εικόνα 64: Ταξινόμηση, όνομα και ομαδοποίηση.

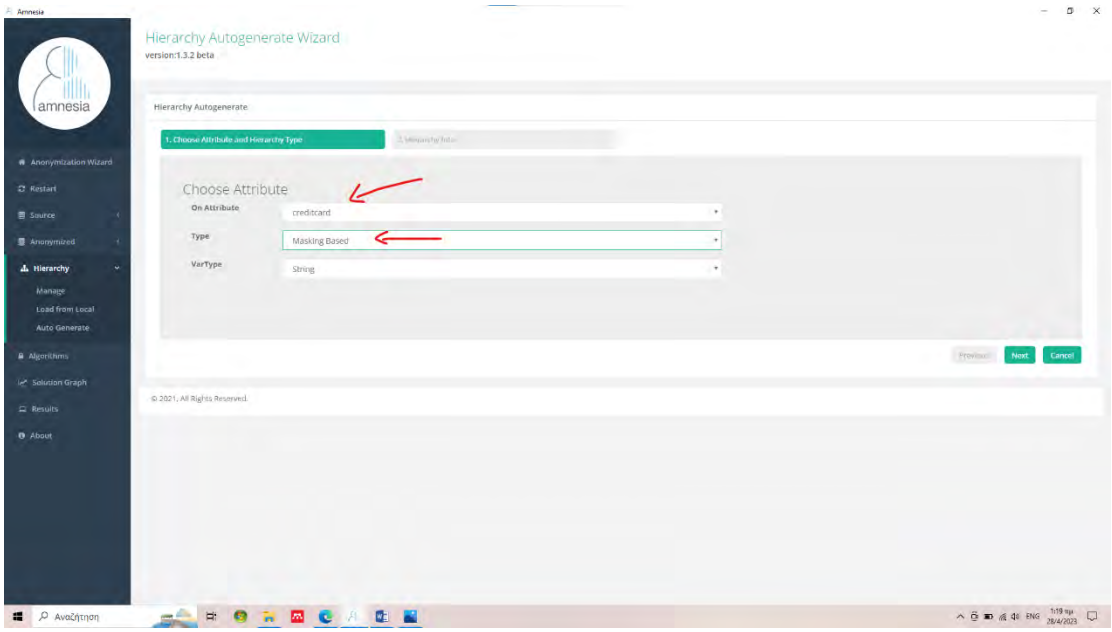
Παρατηρούμε πως ο κανόνας ιεραρχίας έχει δημιουργηθεί, με το προεπιλεγμένο εύρος ομαδοποίησης των δεδομένων μέχρι και το τέταρτο επίπεδο όπου έχουμε την πλήρη γενίκευση του χαρακτηριστικού.



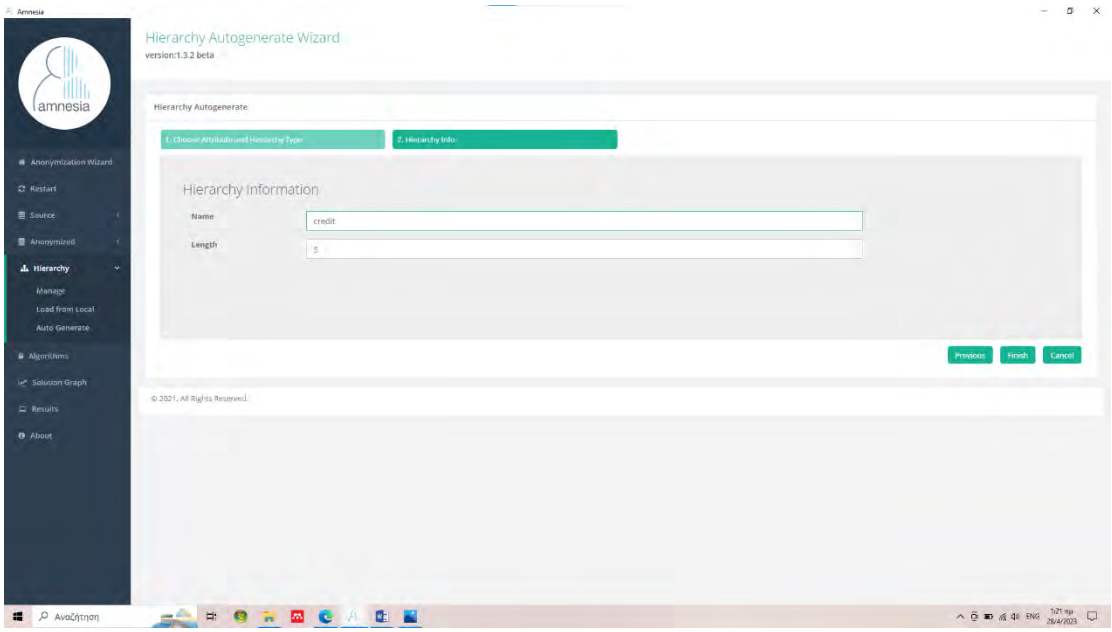
Εικόνα 65: Ομαδοποίηση και γενίκευση.

Ομοίως, για την μεταβλητή zip code οι τιμές 19 έως 1502 γενικεύονται στο 99926, οι τιμές 1514 έως 3586 γενικεύονται στο 99927 κοκ. Τελικά, βλέπουμε πως η ρίζα του δέντρου για τη μεταβλητή zip code με εγγυημένη ανωνυμοποίηση είναι το 100035 που περιλαμβάνει όλες τις τιμές για το zip code.

Ολοκληρώνουμε τους κανόνες ιεραρχίας με το χαρακτηριστικό πιστωτική κάρτα (credit card). Με την ίδια διαδικασία επιλέγουμε Autogenerate Hierarchies και το χαρακτηριστικό credit card. Αυτή τη φορά όμως θα επιλέξουμε στον τύπο χαρακτηριστικού (Type) την επιλογή Masking Based καθώς θα δείξουμε πως μπορούμε να εφαρμόσουμε την τεχνική της συγκάλυψης (masking) μέσω των κανόνων ιεραρχίας. Στη συνέχεια επιλέγουμε το όνομα (Name) και το μάκρος (Length) του χαρακτηριστικού στο οποίο θα ενεργοποιείται η συγκάλυψη, άρα και τα επίπεδα γενίκευσης.

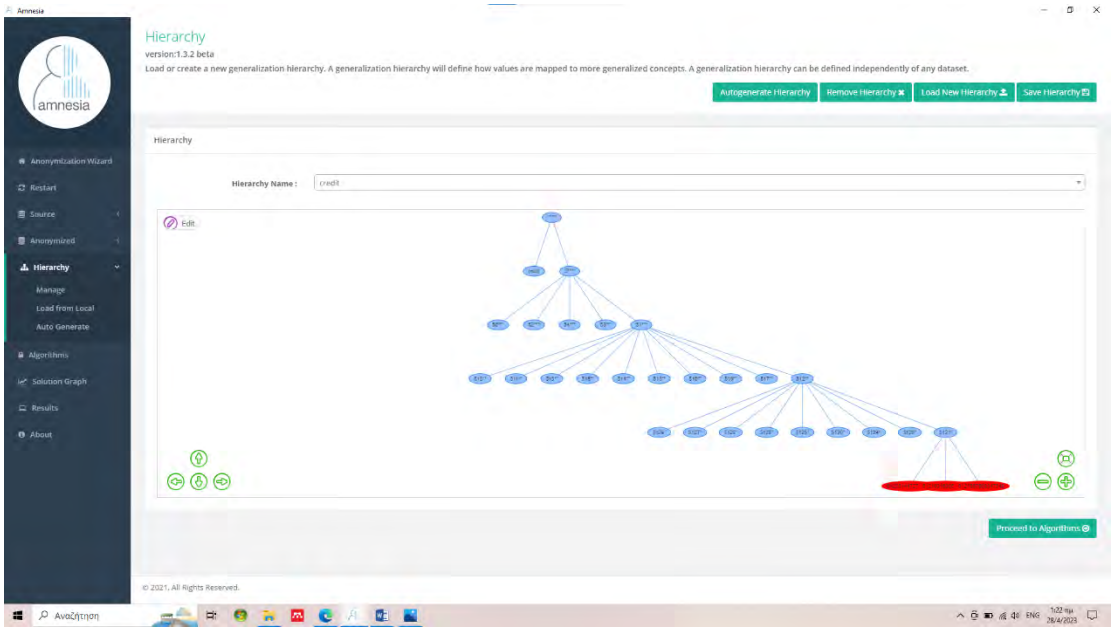


Εικόνα 66:Δημιουργία κανόνα ιεραρχίας για το χαρακτηριστικό πιστωτική κάρτα με χρήση συγκάλυψης. Εικόνα α



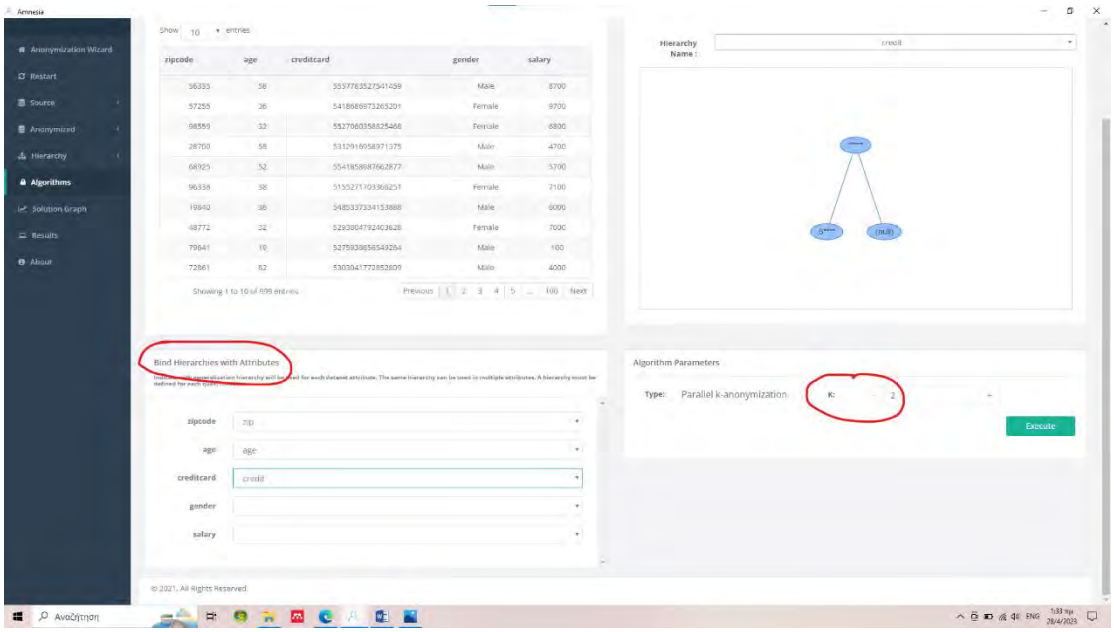
Εικόνα 67: Δημιουργία κανόνα ιεραρχίας για το χαρακτηριστικό πιστωτική κάρτα με χρήση συγκάλυψης. Εικόνα β

Παρατηρούμε πως ολόκληρο το στοιχείο του χαρακτηριστικού εμφανίζεται μόνο στο πρώτο επίπεδο (κόκκινο σύννεφο). Για κάθε επόμενο επίπεδο ξεκινά η συγκάλυψη με αρχή τα πρώτα μπλε σύννεφα όπου υπάρχει συγκάλυψη με ένα χαρακτήρα έως και το τελικό μπλε σύννεφο με συγκάλυψη πέντε χαρακτήρων και πλήρη γενίκευση.



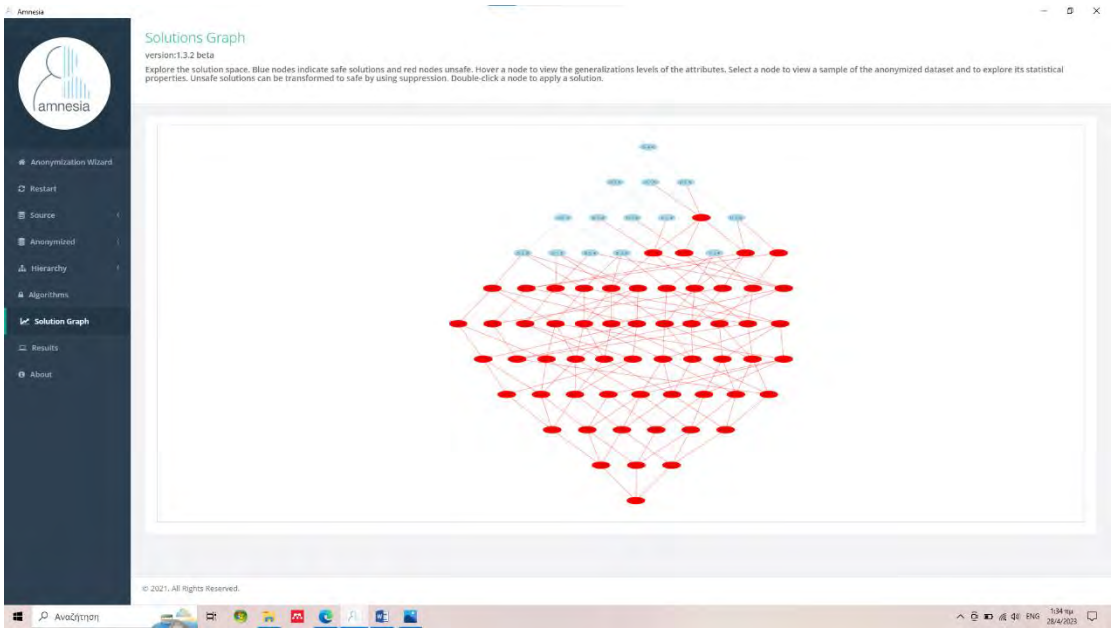
Εικόνα 68: Ομαδοποίηση του χαρακτηριστικού credit card με συγκάλυψη πέντε επιπέδων.

Προχωρώντας στην εφαρμογή αλγόριθμων επιλέγουμε να «δέσουμε» τις ιεραρχίες με τη χρήση του αλγόριθμου k-ανωνυμίας να «δέσουμε» ιεραρχίες με τα χαρακτηριστικά zip, age και credit card, επιλέγοντας για το κάθε χαρακτηριστικό την αντίστοιχη ιεραρχία που έχουμε δημιουργήσει. Επίσης επιλέγουμε k=2 γιατί επιθυμούμε να υπάρξει ανωνυμοποίηση για τουλάχιστον 2 ίδιες εγγραφές οιονεί χαρακτηριστικών στο σύνολο των γενικευμένων δεδομένων μας.

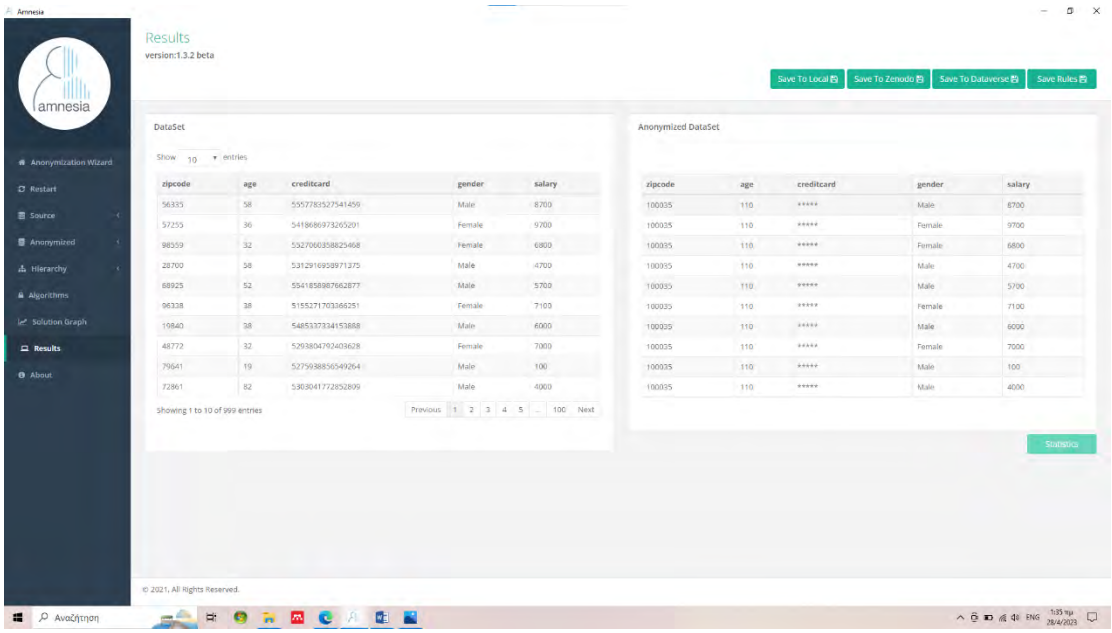


Εικόνα 69: Το "δέσιμο" αλγόριθμων και χαρακτηριστικών.

Όπως έχουμε εξηγήσει (βλ. σελ. 40 του παρόντος) οι κόκκινοι κόμβοι μας υποδεικνύουν μη ασφαλή ανωνυμοποιημένο συνδυασμό γιατί τα δεδομένα έχουν και άλλα επίπεδα γενίκευσης μέχρι να φτάσουμε στην ρίζα του δέντρου. Οι μπλε κόμβοι μας υποδεικνύουν πως όλες οι γενικεύσεις έχουν ολοκληρωθεί με τους κανόνες ιεραρχίας που δημιουργήσαμε και προσφέρουν εγγυημένη ασφάλεια. Σταματώντας τον κέρσορά μας πάνω σε κάθε κόμβο, το λογισμικό είναι σε θέση να μας ενημερώσει το επίπεδο γενίκευσης που έχει επιτευχθεί. Κάνοντας διπλό κλικ στον τελικό κόμβο μας ανοίγει ένα διπλό γραφικό περιβάλλον όπου παρουσιάζονται στο αριστερό και στο δεξί μέρος το σύνολο δεδομένων πριν και μετά την εφαρμογή της ανωνυμοποίησης.



Εικόνα 70: Γράφημα λύσεων. Οι μπλε κόμβοι δίνουν ασφαλείς λύσεις, ενώ οι κόκκινοι μη ασφαλείς. Στη ρίζα του δέντρου βρίσκεται ο κορυφαίος μπλε κόμβος, όπου έχουν επιτευχθεί όλα τα στάδια της γενίκευσης.



Εικόνα 71: Αριστερά το αρχικό σύνολο δεδομένων και δεξιά το ανωνυμοποιημένο σύνολο δεδομένων.

ΚΕΦΑΛΑΙΟ 4

Ανταγωνισμός – Εναλλακτικές εφαρμογές

Στη συνέντευξη που μας παραχώρησε ο κ. Δημακόπουλος, τον ρωτήσαμε αν έχει χρησιμοποιηθεί το λογισμικό Amnesia στην Ελλάδα και στο εξωτερικό, στον ιδιωτικό ή στον δημόσιο τομέα και αν ναι από ποιες εταιρείες. Μας απάντησε ότι δεν μπορεί να γνωρίζει αν έχει χρησιμοποιηθεί ή αν χρησιμοποιείται τώρα αυτή τη στιγμή που μιλάμε σε κάποια ιδιωτική εταιρία ή κάποιο δημόσιο οργανισμό. Μας διαβεβαίωσε ότι έχουν ενδιαφερθεί, έχουν ρωτήσει για το amnesia, έχει παρασχεθεί τεχνική υποστήριξη σε πολλές σκανδιναβικές κυρίως εταιρίες που ασχολούνται με το ασφάλεια δεδομένων και κυβερνοχώρου. Ο κ. Δημακόπουλος μας εξήγησε ότι δεν μπορούσε να μας δώσει περισσότερες πληροφορίες πάνω σε αυτό γιατί υπεύθυνος του τομέα αυτού είναι ο διευθυντής προϊόντος (product manager) του Amnesia. Μας τόνισε πάντως πως κυριότερο ενδιαφέρον έχουμε από εταιρίες που δραστηριοποιούνται σε βόρειες χώρες και στην κεντρική Ευρώπη. Ενδιαφέρονται ιδιαίτερα να προστατεύσουν τους πολίτες τους είτε στον ιδιωτικό τομέα είτε από πανεπιστήμια που μας έχουν ζητήσει από τη Σκανδιναβία. Υπάρχει πολύ μεγάλη κινητικότητα και σε σχολεία όπου μας ζητήθηκε να γίνει παρουσίαση και εργαστήρι (workshop) σε μαθητές ώστε να καταλάβουν τι είναι η ανωνυμοποίηση. Αντίστοιχα στην Ελλάδα μας δήλωσε ότι μόνον εμείς με την παρούσα εργασία και ένας δημοσιογράφος μια φορά έδειξε ενδιαφέρον για το λογισμικό.

Στα πλαίσια της παρούσας έρευνας ρωτήσαμε τον κ. Δημακόπουλο σχετικά με τον ανταγωνισμό που αντιμετωπίζει το λογισμικό Amnesia. Μας δήλωσε πως ο βασικός ανταγωνιστής του Amnesia, σύμφωνα με τον υπεύθυνο προϊόντος (product manager), είναι το εργαλείο ARX. Συνέχισε πως και εκεί πέρα έχει γίνει πολύ σοβαρή δουλειά, τονίζοντας πως είναι το εργαλείο που έχουνε μελετήσει πιο έντονα. Έχει γίνει πολύ σοβαρή δουλειά από μία μεγάλη ομάδα, έχει πολλά concepts, ο χρήστης μπορεί να κάνει πολλά πράγματα, αλλά θεώρησε ότι δεν είναι εύχρηστο. Δηλαδή παρέθεσε πως και αυτός που ασχολείται με την ανωνυμοποίηση πολλές φορές δυσκολεύεται για να βρει κάτι μέσα σε αυτό το εργαλείο. Δεν είναι τόσο εύχρηστο για τον απλό χρήστη. Δηλαδή να μην ξεχνάμε ότι θέλουμε να απευθυνθούμε και σε απλούς πολίτες, όπως παραδείγματος χάριν να μπορεί ο καθένας να ανωνυμοποιήσει τα δεδομένα του. Μπορεί ο οποιοσδήποτε να θέλει να δημοσιοποιήσει ένα σύνολο δεδομένων του για στατιστικούς λόγους. Οπότε είπε χαρακτηριστικά ο κ Δημακόπουλος ότι εμείς απευθυνόμαστε και στον απλό πολίτη τον καθημερινό, να είναι σε θέση να κάνει μια ανωνυμοποίηση. Τέλος παραδέχθηκε χωρίς αντιρρήσεις πως αναγνωρίζει

ότι είναι ένα δύσκολο εγχείρημα αυτό που επιδιώκει η ομάδα του Amnesia, αλλά πρέπει να προσπαθήσουν να κάνουνε ακόμη πιο απλά τα πράγματα για να μπορέσουνε να διευκολύνουνε τους χρήστες του λογισμικού.

4.1 ARX

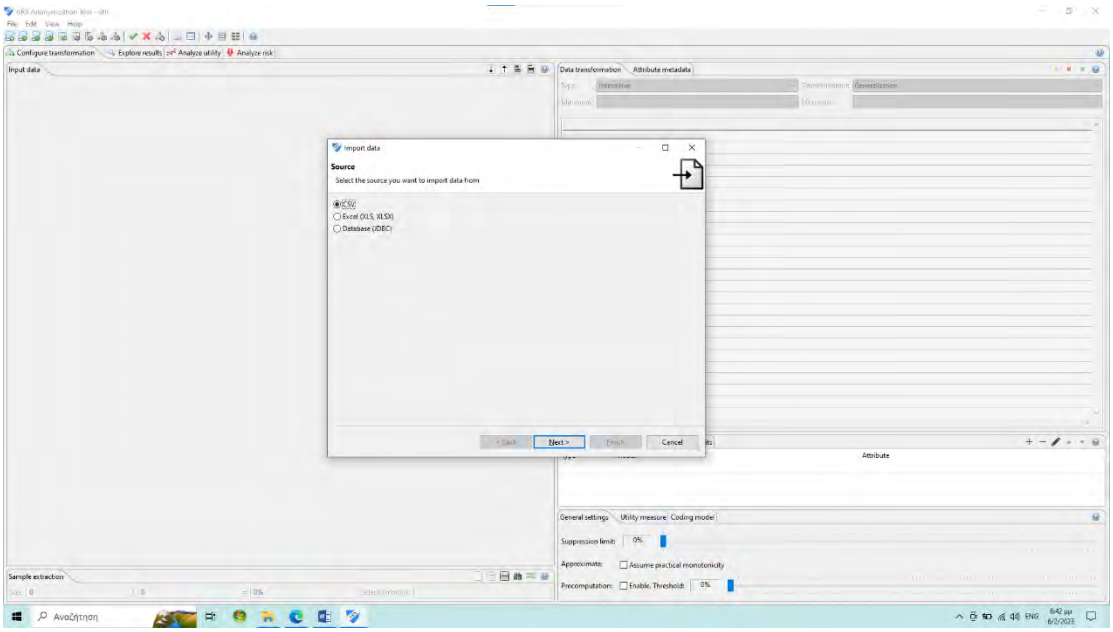
Η εφαρμογή ARX είναι ένα λογισμικό ανοιχτού κώδικα (open source) που παρέχεται δωρεάν. Μέσα από την ιστοσελίδα της εταιρίας https://arx.deidentifier.org στην κατηγορία downloads μπορούμε να κατεβάσουμε το εκτελέσιμο αρχείο. Η διαδικασία εγκατάστασης είναι πολύ εύκολη, όπως μία οποιαδήποτε εγκατάσταση.(ARX - Data Anonymization Tool web page, last accessed 03/04/2023)

Το ARX αποτελεί την πληρέστερη εφαρμογή ανωνυμοποίησης δεδομένων. Είναι μία από εφαρμογές που παρέχει τις περισσότερες λειτουργείες και ενεργή υποστήριξη. Ένα λογισμικό που εισαγάγει πληθώρα πρωτοκόλλων ασφαλείας με ιδιαίτερα φιλικό τρόπο, προσφέρει άμεσο πολύ λειτουργικό γραφικό περιβάλλον, τα αποτελέσματα των διαδικασιών του μπορούν να ενσωματωθούν σε άλλα λογισμικά. (Prasser et al., 2014)

Αναπτύχθηκε από το Ινστιτούτο Υγείας του Βερολίνου στη Γερμανία και δημοσιεύτηκε για πρώτη φορά το 2012. Εστιάζει στην πλήρη αυτοματοποίησης και ευελιξία. Υποστηρίζει ένα ευρύ φάσμα μοντέλων απορρήτου, μοντέλων μετασχηματισμού και μοντέλων χρησιμότητας που μπορούν να συνδυαστούν χωρίς περιορισμό. Αυτή η ευελιξία επιτυγχάνεται χάρη σε ένα κεντρικό - γενικό αλγόριθμο πυρήνα, σε συνδυασμό με ένα φιλικό περιβάλλον χρόνου εκτέλεσης προσαρμοσμένο στις εργασίες ανωνυμοποίησης που έχει υιοθετηθεί από πολλά άλλα συστήματα, όπως το Amnesia. Το ARX είναι διαθέσιμο ως λογισμικό ανοιχτού κώδικα, εναρμονισμένο με όλα τα μεγάλα λειτουργικά συστήματα και είναι ακόμα υπό ενεργό ανάπτυξη με τελευταία ενημέρωση το 2022. (Haber et al., 2022)

Θέλοντας να αναφερθούμε στα θετικά από τη σκοπιά του ερευνητή, η εφαρμογή υποστηρίζει:

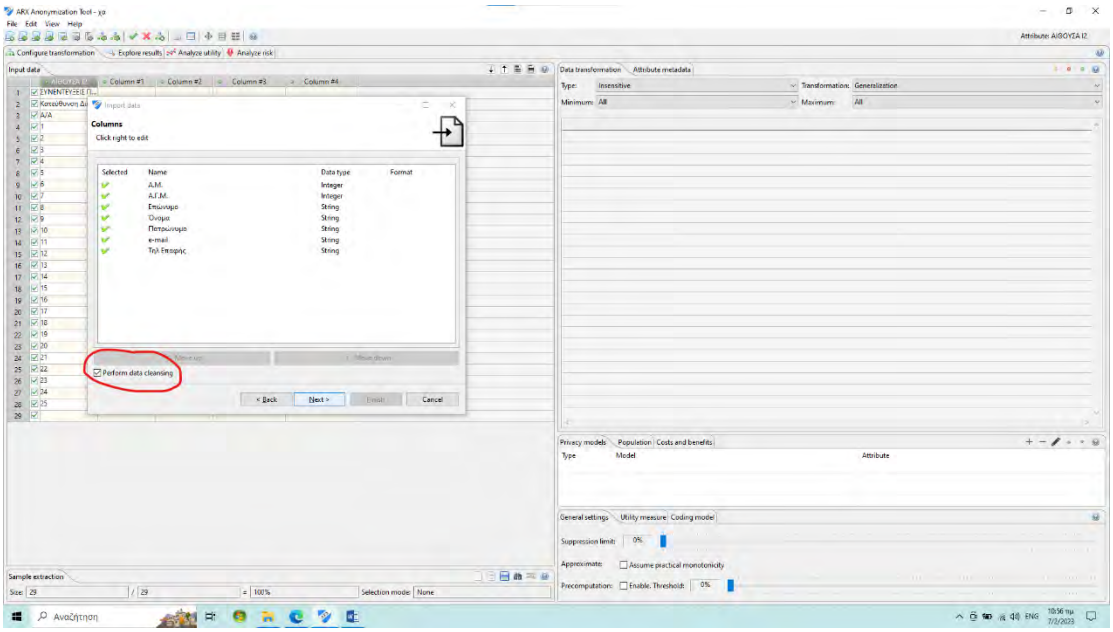
- όλα τα διαθέσιμα αρχεία για ανωνυμοποίηση όπως .csv, .xls ή αρχεία Database



Εικόνα 72: Εισαγωγή δεδομένων στο λογισμικό ARX.

- όλες τις πρόσφορες γλώσσες

- αναγνωρίζει αυτόματα όλα τα είδη και τις κατηγορίες δεδομένων του επιλεγμένου αρχείου μας, ενώ εξ' αρχής δίνει τη δυνατότητα καθαρισμού τυχόν δεδομένων που δεν χρειάζονται



Εικόνα 73: Έλεγχος και καθαρισμός δεδομένων από το λογισμικό ARX.

- μας παρέχει για κάθε κατηγορία του συνόλου δεδομένων μας προς ανωνυμοποίηση, ξεχωριστή διαδικασία και πρωτόκολλο ασφαλείας, αντί για μια γενικευμένη ανωνυμοποίηση όλων των δεδομένων.

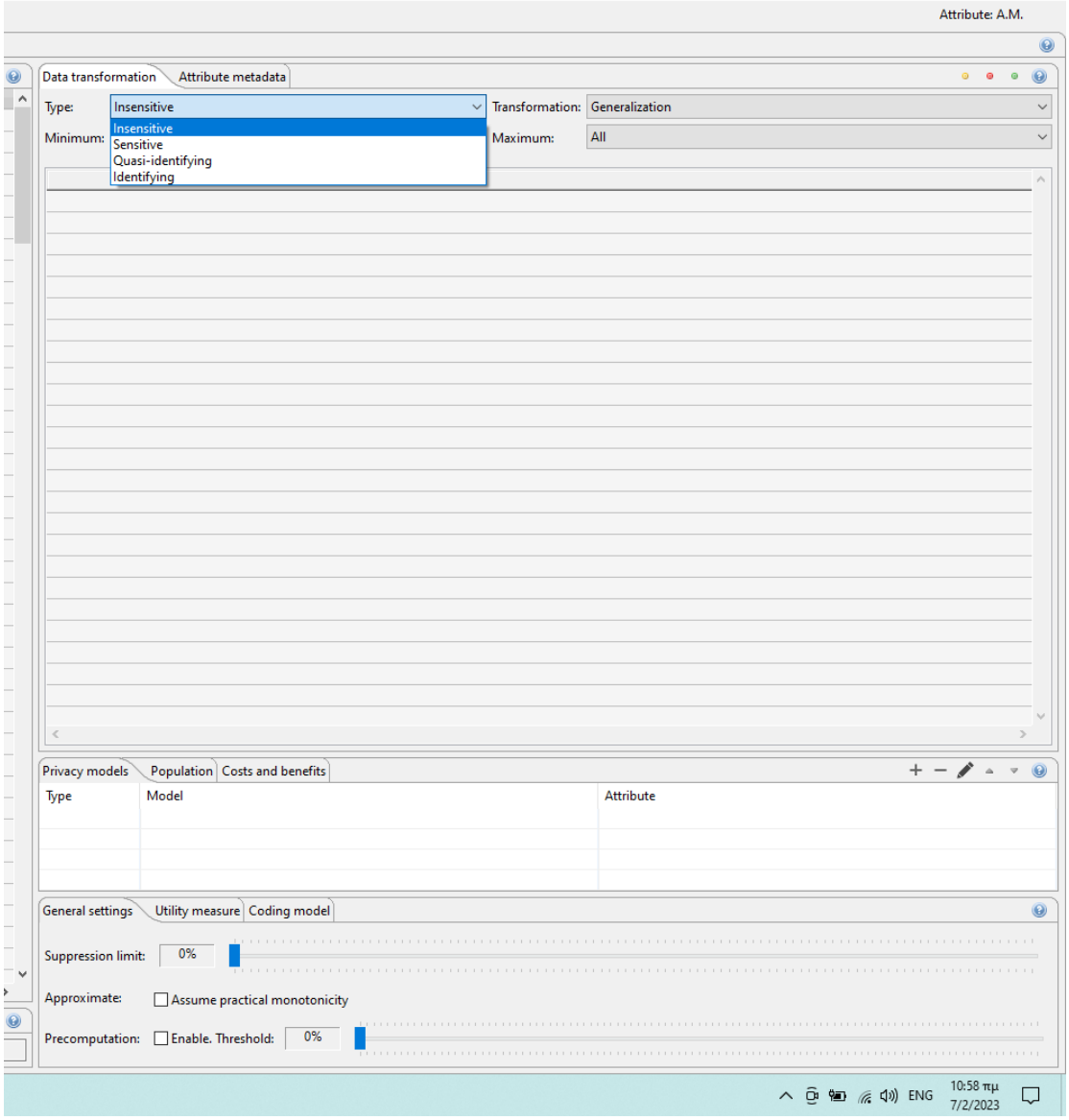
- Μας δίνει επίσης τη δυνατότητα να ξεχωρίσουμε τα δεδομένα σε:

α) Μη ευαίσθητα (insensitive) – χωρίς τροποποίηση

β) Ευαίσθητα (sensitive) – αυτά που θα ανωνυμοποιηθούν

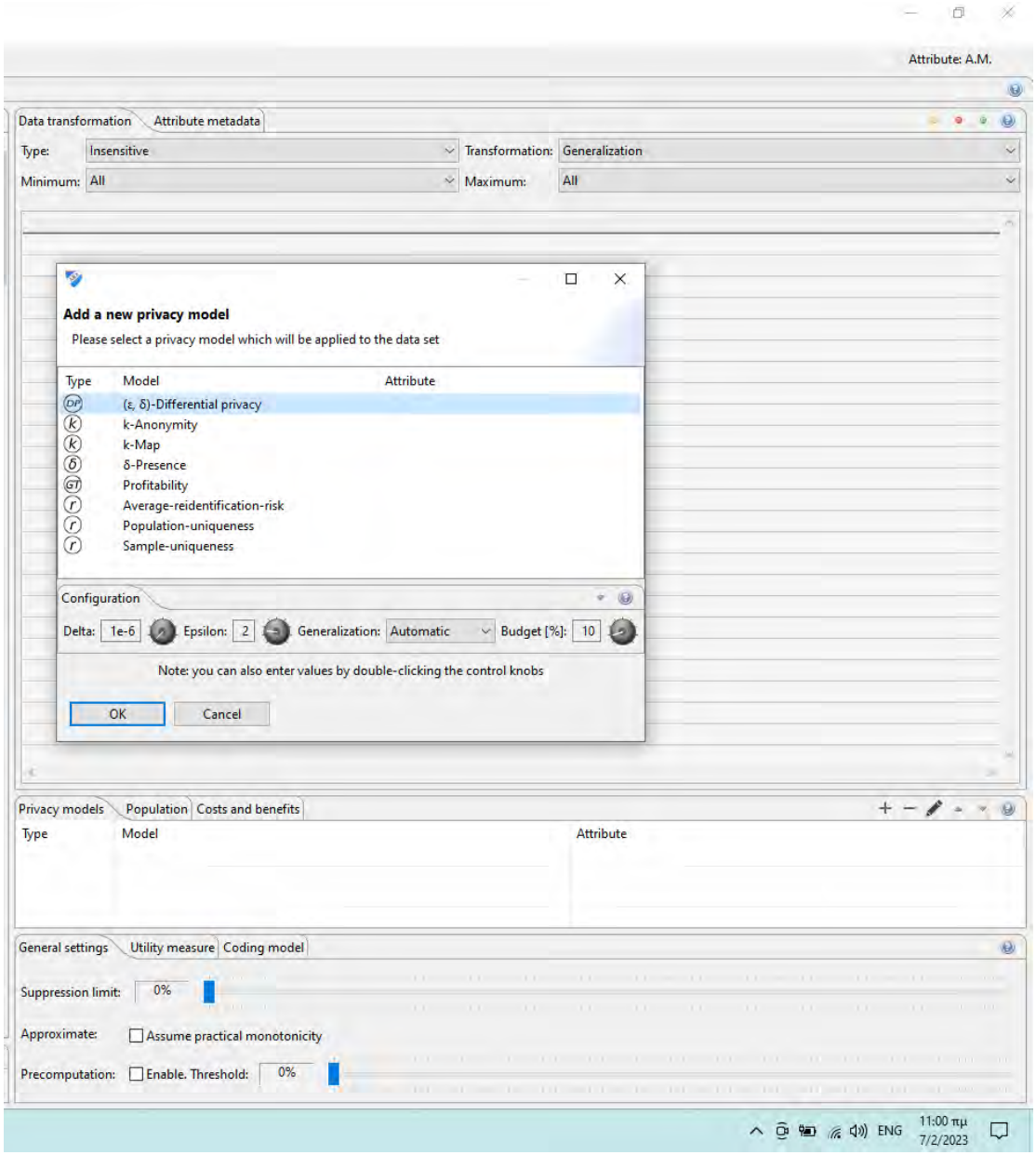
γ) Οιονεί προσδιοριστικό (quasi-identifying) – αυτά που θα τροποποιηθούν

δ) Αναγνώρισης (Identifying) – αυτά που θα αφαιρεθούν από το σύνολο δεδομένων



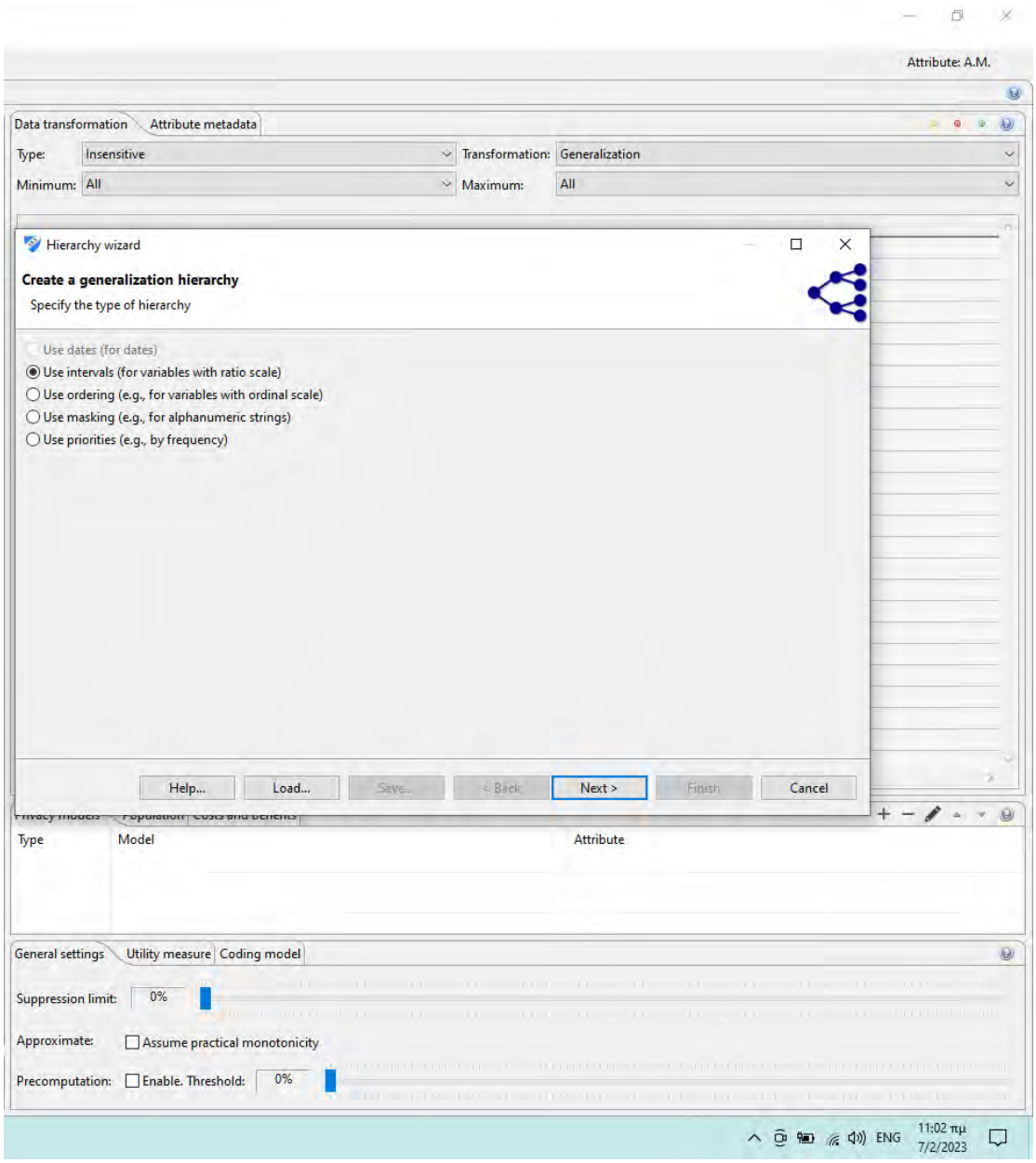
Εικόνα 74: Επιλογές χαρακτηριστικών. ARX

- Χρησιμοποιεί όλα τα διαθέσιμα πρωτόκολλα ασφαλείας για κάθε κατηγορία δεδομένων χωριστά όπως διαφορικό απόρρητο, k-ανωνυμία κλπ



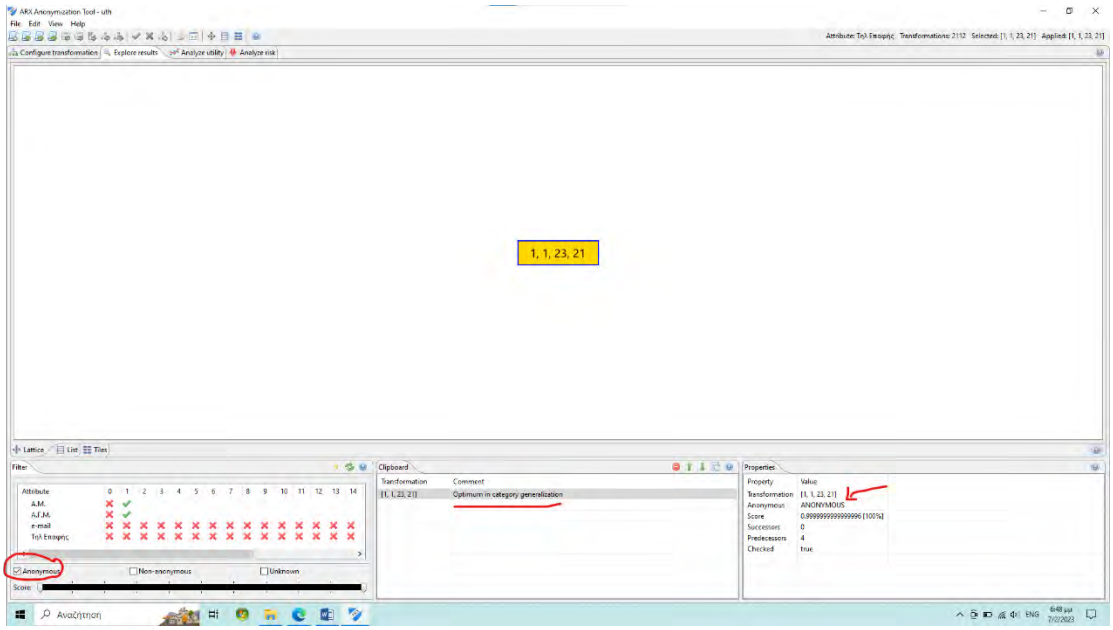
Εικόνα 75: Επιλογές αλγόριθμων. ARX

- Χρησιμοποιεί κανόνες ιεραρχίας για κάθε κατηγορία δεδομένων χωριστά είτε επιλέγοντάς τους αυτόματα με βάση τα δεδομένα που έχουμε ή επιτρέπει την χειροκίνητη επιλογή

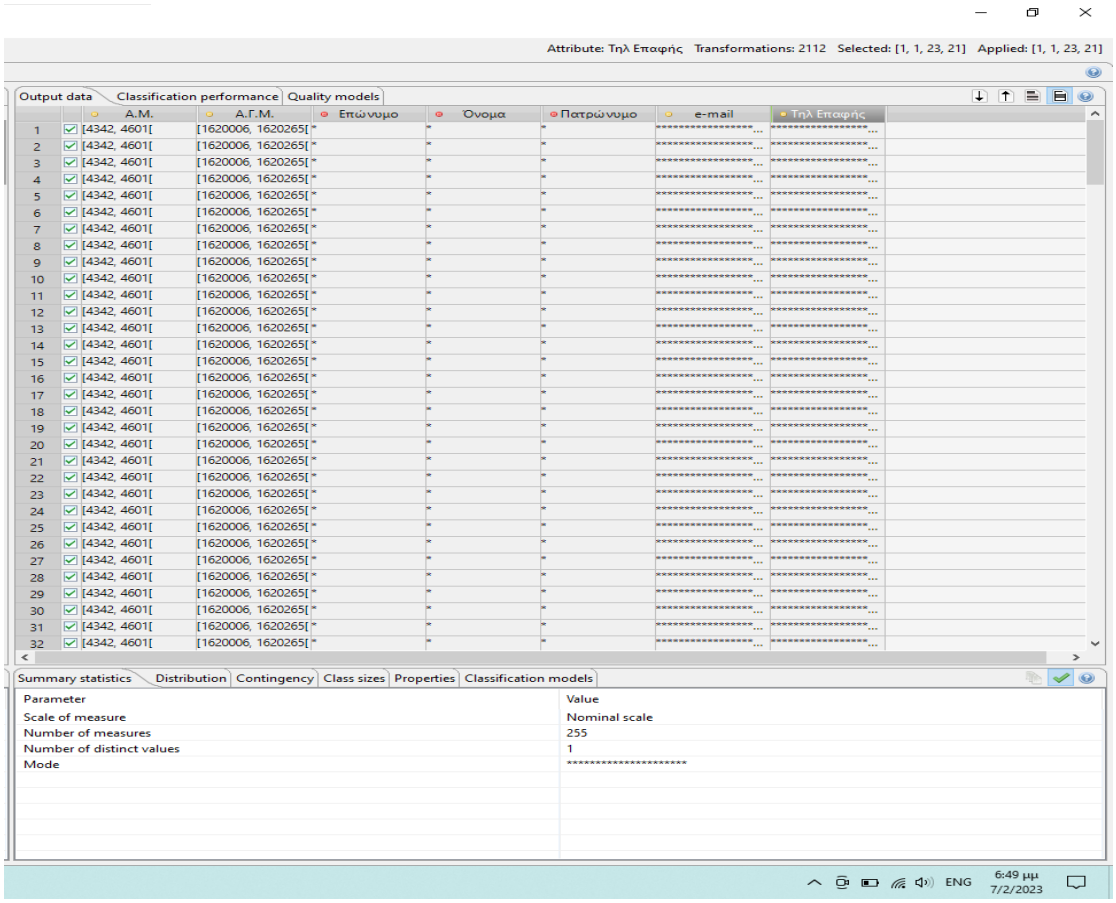


Εικόνα 76: Επιλογές κανόνων ιεράρχησης. ARX

- Μας δίνει την δυνατότητα σε κάθε στάδιο της διαδικασίας ανωνυμοποίησης ή επιλογής κανόνων ιεράρχησης, την αλλαγή των επιλογών που έχουμε κάνει, χωρίς καμία δέσμευση για το τελικό αποτέλεσμα, μας ενημερώνει για τις αλλαγές που έχουν πραγματοποιηθεί και αν έχει ανωνυμοποιηθεί το σύνολο δεδομένων μας.



Εικόνα 77: Έλεγχος και προεπισκόπηση δεδομένων. ARX



Εικόνα 78: Απεικόνιση ανωνυμοποιημένων δεδομένων με το ARX.

- Τέλος μπορούμε να εξάγουμε το ανωνυμοποιημένο σύνολο δεδομένων σε .csv αρχείο.

Θέλοντας να αναφερθούμε στα αρνητικά της εφαρμογής από τη σκοπιά του ερευνητή:

- Παρόλο που μπορούμε να επιτρέψουμε την εφαρμογή να εκτελέσει τα πάντα αυτόματα, εντούτοις χρειάζεται αρκετή εξοικείωση από τον υποψήφιο χρήστη για τις εξατομικευμένες επιλογές
- Παρέχει αρκετά μεγάλο όγκο πληροφορίας, ο οποίος για μέσο χρήστη μπορεί να φανεί λίγο κουραστικός και δύσκολος στην κατανόηση συγκριτικά με άλλες εφαρμογές
- Εξάγει τα ανωνυμοποιημένα σύνολα δεδομένων μόνο σε .csv αρχεία
- Δεν χρησιμοποιεί καθόλου τον αλγόριθμο της km-ανωνυμίας
- Δεν παρέχει αυτόματη διαδικασία ανωνυμοποίησης με καθοδήγηση του χρήστη όπως το Amnesia

4.2 BIZDATA-X

Η εφαρμογή BIZ-DATA-X είναι μία εφαρμογή που παρέχεται επί πληρωμή. Δίνεται η δυνατότητα για μια δοκιμαστική έκδοση 30 ημερών έπειτα από μία απλή ηλεκτρονική αίτηση στην κατασκευάστρια εταιρία.

Η εγκατάσταση δεν ήταν μία απλή διαδικασία διότι για την ολοκλήρωσή της χρειάζονταν εγκατάσταση πολλών πρόσθετων χαρακτηριστικών – εφαρμογών. Πιο συγκεκριμένα χρειάζονταν πρόσθετο της Microsoft (SCCM 2007) και του πρόσθετου webdav.

Σε οποιαδήποτε προσπάθεια εγκατάστασης παρουσίαζε σφάλμα.

Σε προσπάθεια επικοινωνίας με την κατασκευάστρια εταιρία δεν λάβαμε ποτέ απάντηση για το πρόβλημα εγκατάστασης που αντιμετωπίσαμε.

4.3 G9 Anonymizer

Η εφαρμογή G9-Anonymizer είναι μία εφαρμογή δυσκολότερη στην εγκατάστασή της από μία τυπική εγκατάσταση εφαρμογής. Για να μπορέσουμε να τη χρησιμοποιήσουμε έπρεπε να εγκαταστήσουμε πρώτα μια άλλη εφαρμογή (eclipse marketplace) και έπειτα το G9 anonymizer.

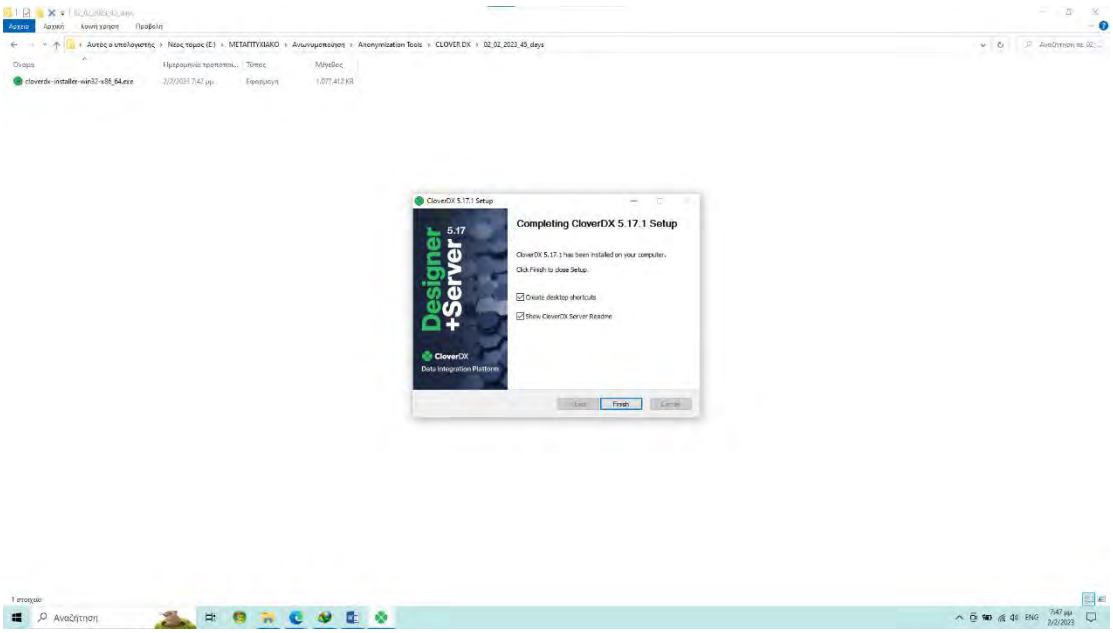
Η εφαρμογή είναι επί πληρωμή. Παρόλο που διαφημίζει η κατασκευάστρια εταιρία του λογισμικού δωρεάν δοκιμαστική έκδοση για επτά (7) ημέρες, εντούτοις δεν μας παρείχε αυτή τη δυνατότητα με αποτέλεσμα να μην μπορέσουμε να τη δοκιμάσουμε. Όποια προσπάθεια

επικοινωνίας με την κατασκευάστρια εταιρία του λογισμικού απέβει άκαρπη, καθώς ουδέποτε ανταποκρίθηκε στα μηνύματά μας.

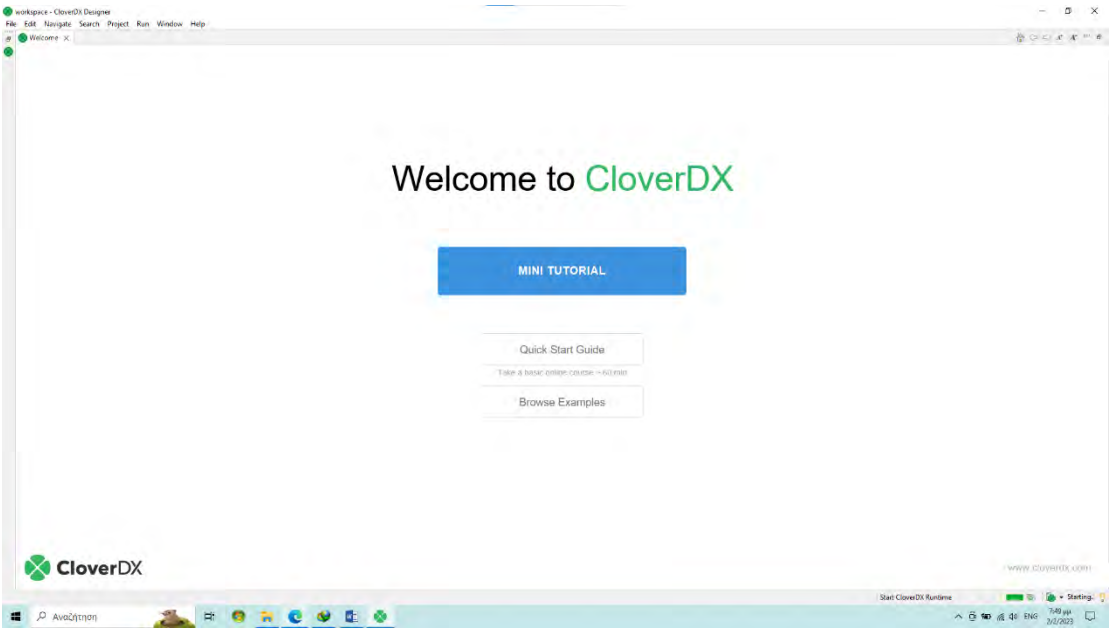
4.4 Clover DX

Η εφαρμογή CLOVER DX είναι μία εφαρμογή που παρέχεται επί πληρωμή. Δίνεται η δυνατότητα για μια δοκιμαστική έκδοση σαράντα πέντε (45) ημερών έπειτα από μία ηλεκτρονική αίτηση την κατασκευάστρια εταιρία. Το αρχείο εγκατάστασης ήταν μεγαλύτερο από τα συνηθισμένα αρχεία προς εγκατάσταση 1,02GB.

Πολύ βοηθητική εφαρμογή καθώς από την αρχική οθόνη μας δίνεται η δυνατότητα για μικρό οδηγό (mini tutorial) ή για σύντομο οδηγό χρήσης (quick Start Guide).

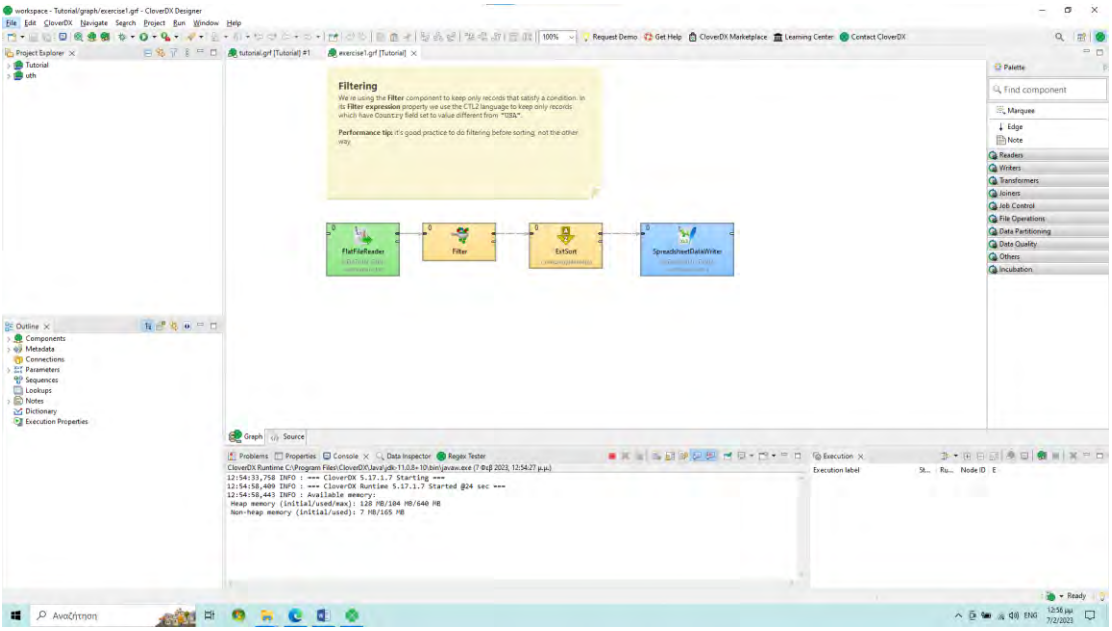


Εικόνα 79: Εγκατάσταση cloverDX



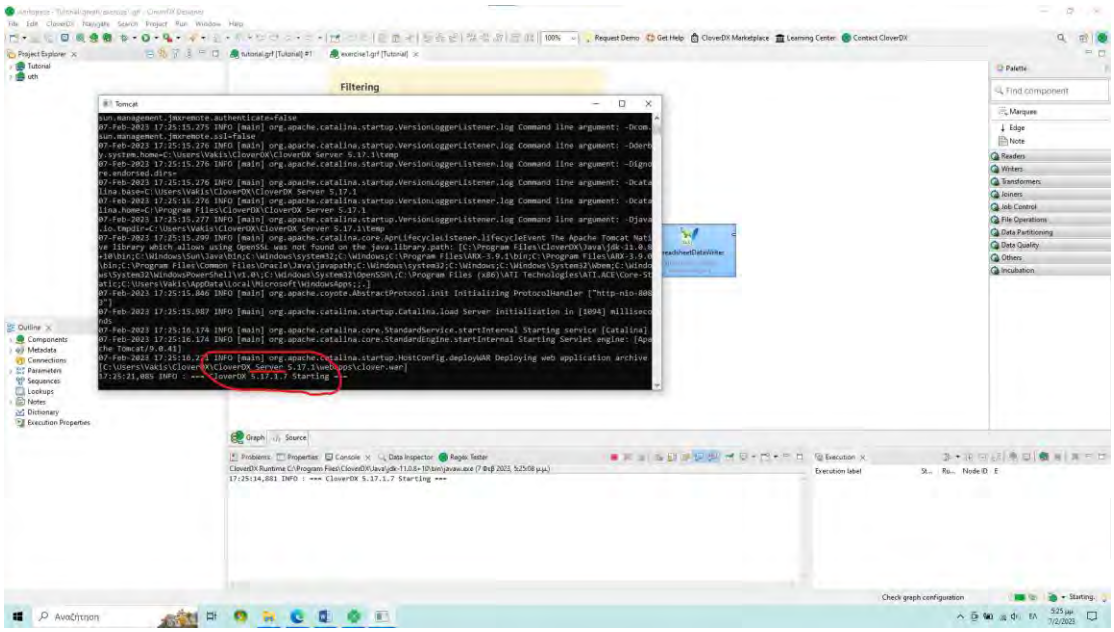
Εικόνα 80: Κεντρικό παράθυρο λογισμικού.

Η εφαρμογή παρέχει πάρα πολλές λειτουργίες οι οποίες μπορούν σε κάθε φάση της επεξεργασίας των δεδομένων να τις παρακολουθεί και να επεμβαίνει ο χρήστης της. Χρειάζεται όμως περισσότερες γνώσης πληροφορικής από αυτές ενός απλού χρήστη καθώς χρησιμοποιεί γραμμικό κώδικα για την λειτουργία της ή την επεξεργασία των δεδομένων, γεγονός που το καθιστά δύσχρηστο στο ευρύ κοινό.



Εικόνα 81: Διαδικασία ανωνυμοποίησης με το cloverDX.

Αν και όλες οι εντολές δίνονται μέσω του εγκατεστημένου προγράμματος, οι εκτελεσθείσες αλλαγές πραγματοποιούνται μέσω της λειτουργίας clover DX server. Επίσης μέσω του clover DX server μπορούν όλοι οι συνδεδεμένοι χρήστες να το παρακολουθούν, επεμβαίνοντας σε όλα τα σενάρια και σε όλα τα στάδια. Μπορούν επίσης να διαχειρίζονται τα διάφορα συμβάντα, να παραμετροποιούν και να διαχειρίζονται όλα τα λάθη στη διαδικασία της ανωνυμοποίησης ενός συνόλου δεδομένων. Επίσης μπορεί να λειτουργήσει σενάρια ανωνυμοποίησης ατομικά με ένα χρήστη σε ένα ηλεκτρονικό υπολογιστή ή να διαμοιράσει όλα τα σενάρια σε πολλαπλούς χρήστες και ηλεκτρονικούς υπολογιστές.

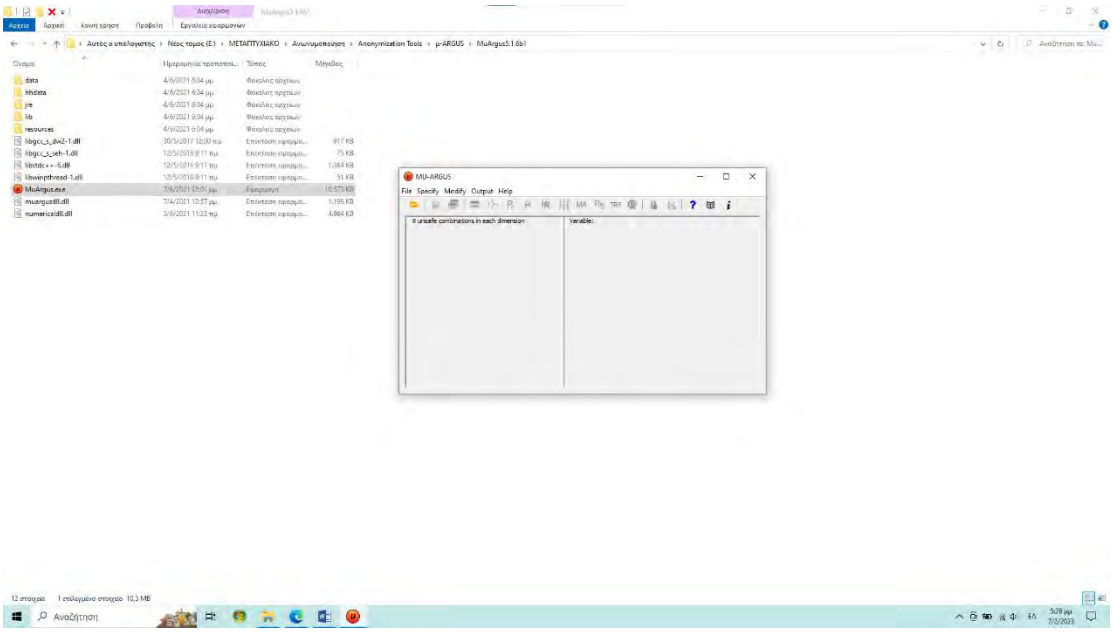


Εικόνα 82: Απεικόνιση εργασίας με CloverDX server

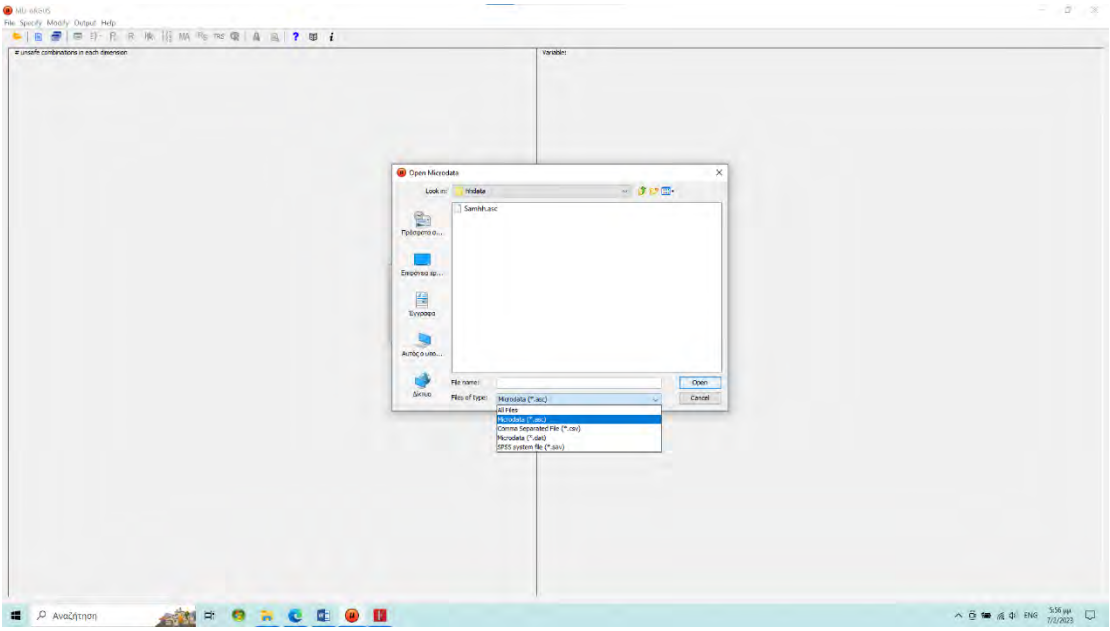
Σε περίπτωση χρονοδιαγράμματος υπάρχει ομάδα υποστήριξης που μπορεί να επέμβει βοηθητικά μέχρι του σημείου που θα υποδείξουμε, ώστε να μας βοηθήσει να τελειώσουμε ένα σενάριο της υπηρεσίας μας ή της επιχείρησής μας.

4.5 μ-ARGUS

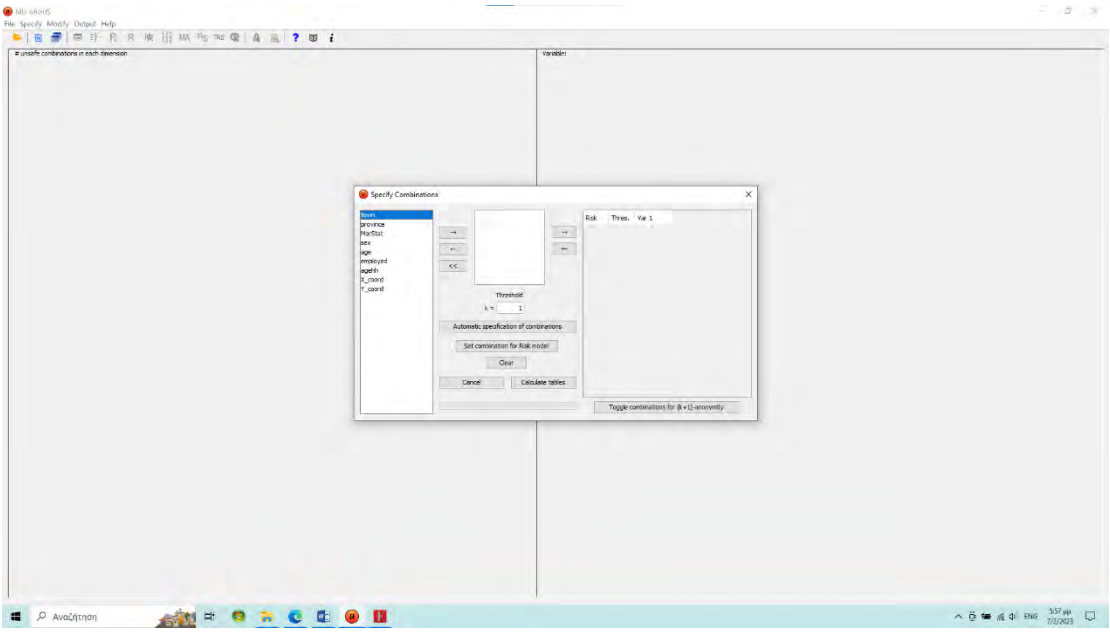
Μία εφαρμογή ανωνυμοποίησης που δοκιμάσαμε είναι η εφαρμογή μ-ARGUS. Το μεγαλύτερο πλεονέκτημα αυτής της εφαρμογής είναι ότι λειτουργεί χωρίς την ανάγκη εγκατάστασης.



Εικόνα 83: Το λογισμικό μ-Argus.



Εικόνα 84: Εισαγωγή αρχείου συνόλου δεδομένων μ-Argus.



Εικόνα 85: Επιλογή συνδυασμών m-Argus.

Έχει ενσωματωμένο εγχειρίδιο οδηγιών και έτοιμα σύνολα δεδομένων για πρακτική εξάσκηση. Υποστηρίζει πληθώρα επιλογών και συνδυασμών, όμως χρησιμοποιεί μόνο ως πηγή, σύνολα δεδομένων, αρχεία μικροδεδομένων (microdata).

Η χρήση αρχείων μικροδεδομένων όπως έχουμε αναφέρει στην ενότητα 1.12 απαιτεί άδεια και τα πλαίσια έρευνας είναι πιο στενά και συγκεκριμένα, ενώ παρέχεται μεγαλύτερη ευκολία με σύνολα δεδομένων του Excel.

ΚΕΦΑΛΑΙΟ 5

5.1 Πλεονεκτήματα του λογισμικού Amnesia

Σύμφωνα με τη συνέντευξη που μας παραχώρησε ο κ. Δημακόπουλος παραδέχτηκε ότι υπάρχει πολύ επιτυχημένη ανατροφοδότηση (feedback) σχετικά με τη φιλικότητα του περιβάλλοντος χρήσης (UI-User Interface) του Amnesia. Χαρακτηριστικά είπε: *«Τους έχει κερδίσει το UI»*.

Η ανωνυμοποίηση δεν είναι μια απλή διαδικασία καθώς πραγματοποιούνται ρυθμίσεις και διαδικασίες που απαιτούν επεξήγηση και ανάλυση για κάθε χρήστη, ειδικά στο μέσο χρήστη, που θα έρθει να ανωνυμοποιήσει για πρώτη φορά ένα σύνολο δεδομένων. Το λογισμικό εξαιτίας της εξαιρετικής φιλικότητάς του προς τον χρήστη, μέσω του περιβάλλοντος χρήσης, καθοδηγεί αρκετά και το καθιστά ευκολότερο και λειτουργικότερο από άλλα εργαλεία ανωνυμοποίησης.

Στα πλεονέκτημα του Amnesia είναι η «ελαφριά» κατασκευή του. Δηλαδή είναι ένα λογισμικό το οποίο μπορεί να εκτελέσει τις λειτουργίες του τόσο διαδικτυακά μέσω της Online επιλογής, όσο και εγκαθιστώντας το σε οποιοδήποτε υπολογιστή, αφού είναι σχεδιασμένο με βάση την Java, δεν απαιτείται υψηλή υπολογιστική ισχύ και η χρήση του μπορεί να πραγματοποιηθεί ακόμη και σε παλαιότερης τεχνολογίας ηλεκτρονικούς υπολογιστές.

Το Amnesia μπορεί να εφαρμόσει πλήρη και τοπική γενίκευση ανά χαρακτηριστικό και να ενσωματώσει προσαρμοσμένες μεθόδους στα σύνολα δεδομένων, δηλαδή μπορεί ο χρήστης να επέμβει και να αλλάξει τον τύπο χαρακτηριστικού. Σημαντικό είναι να αναφερθεί πως το λογισμικό Amnesia είναι το μοναδικό που εφαρμόζει την km - ανωνυμία, το οποίο είναι ένα μοντέλο απορρήτου που μεταφέρει την έννοια της k-ανωνυμίας σε δεδομένα συνόλου.

Με τη δωρεάν χρήση του ως λογισμικό ανοιχτού κώδικα, δίνεται η δυνατότητα να χρησιμοποιηθεί από οποιονδήποτε με σκοπό τη δοκιμή η την έρευνα, είτε πρόκειται να χρησιμοποιηθεί για τον σκοπό για τον οποίο δημιουργήθηκε δηλαδή την ανωνυμοποίηση συνόλων δεδομένων.

Θετικό στοιχείο επίσης του Amnesia από την εμπειρία του ερευνητή, ήταν ένα πρόβλημα συμβατότητας του προγράμματος με το πρόσθετο java, το οποίο όμως αναφέρθηκε και αντιμετωπίστηκε - επιλύθηκε με μεγάλη ταχύτητα και αμεσότητα.

5.2 Μειονεκτήματα του λογισμικού

Ένα από τα μειονεκτήματα του Amnesia είναι η «ηλικία» του. Το λογισμικό είναι εξαιρετικά καινούριο στην αγορά (2019) και σε αντίθεση με άλλα λογισμικά (πχ ARX) τα οποία υπάρχουν εδώ και είκοσι χρόνια είναι απολύτως κατανοητό ότι δεν μπορεί να έχει τον ίδιο όγκο πάνω σε αλγόριθμους και μοντέλα ανωνυμοποίησης. Στη συνέντευξη που μας παραχώρησε ο κ. Δημακόπουλος, μας ανέφερε πως σύντομα στην επόμενη έκδοση του λογισμικού θα υπάρξει ένας νέος αλγόριθμος μια εύκολη εφαρμογή του διαφορικού απορρήτου (differential privacy) πάνω στα σύνολα δεδομένων (dataset), που ουσιαστικά αυτό το σχέδιο (concept) θα πραγματοποιεί τη μικρότερη απώλεια πληροφορίας (information loss).

Ένα άλλο μειονέκτημα είναι πως επειδή είναι σε ερευνητικό επίπεδο, μπορεί να υπάρξουν τεχνικά προβλήματα και δυσλειτουργίες (bug). Γι' αυτό προτείνεται στους χρήστες αν εντοπίσουν κάποιο πρόβλημα, να το αναφέρουν στην πλατφόρμα επικοινωνίας, ώστε να μπορέσουν οι διαχειριστές να το επιλύσουν – διορθώσουν.

5.3 SWOT ανάλυση του λογισμικού

Η SWOT ανάλυση είναι ένα απλό και πολύ συχνά χρησιμοποιούμενο εργαλείο για να μπορέσουμε να συγκρίνουμε τα δυνατά και τα αδύναμα σημεία στο εσωτερικό περιβάλλον, με τις ευκαιρίες και τις απειλές στο εξωτερικό περιβάλλον, ενός οργανισμού ή μιας επιχείρησης, ενός έργου ή ακόμη και ενός κλάδου ή επιχείρησης. Το SWOT σημαίνει τα δυνατά (Strengths), τα αδύναμα (Weaknesses), οι ευκαιρίες (Opportunities) και οι απειλές (Threats) για την κάθε ξεχωριστή περίπτωση. Στην περίπτωσή μας το εσωτερικό και το εξωτερικό περιβάλλον του λογισμικού. Η ανάλυση SWOT είναι εξαιρετικά ευέλικτη και μπορεί να χρησιμοποιηθεί σε οποιοδήποτε επίπεδο ανάλυσης και μας δείχνει την κατάσταση που επικρατεί τη στιγμή που πραγματοποιήθηκε η ανάλυση (Teece, 2018).

S 1. Φιλικό περιβάλλον χρήστη (UI) 2. Γραφικό περιβάλλον των λύσεων του κάθε σεναρίου 3. Αλγόριθμος Km-Ανωνυμία 4. Ελαφριά κατασκευή χωρίς απαιτήσεις σε επεξεργαστική ισχύ 5. Διατίθεται δωρεάν	**W** 1. Πρόσφατο λογισμικό 2. Λιγότερες λύσεις ανωνυμοποίησης 3. Δυσκολία διατήρησης αν παραμείνει σε ερευνητικό επίπεδο και εξαντληθεί η χρηματοδότηση
O 1. Εισαγωγή νέων αλγόριθμων σε επόμενη αναβάθμιση 2. Δημιουργία εταιρικού προφίλ και εισαγωγή στην παγκόσμια αγορά 3. Επίσημη επί πληρωμή έκδοση του λογισμικού	**T** 1. Το λογισμικό ARX, καθώς ο σχεδιασμός και η μελέτη του Amnesia στηρίζονται σε αυτό 2. Πλήθος εναλλακτικών εφαρμογών 3. Λειτουργία των οργανισμών με ειδικευμένα λογισμικά του κλάδου πχ. Πλατφόρμα my school, κτηματολόγιο ΑΕ κλπ

Εικόνα 86: Μία SWOT ανάλυση του λογισμικού Amnesia

5.4 Μελλοντικά σχέδια για το λογισμικό Amnesia

Σε συνέχεια της συνέντευξης μας (10ος 2022), ρωτήσαμε τον κ. Δημακόπουλο σχετικά με τα μελλοντικά σχέδια για το Amnesia, ειδικά όταν μπορούν οι εταιρίες ή οι οργανισμοί να λειτουργήσουν σε ένα κλειστό εσωτερικό δίκτυο ή vpn ή κρυφή ip μέσα σε μια εταιρία ή έναν οργανισμό. Η απάντησή του ήταν ότι μπορεί οι περισσότερες εταιρίες να χρησιμοποιούν όλα αυτά που προαναφέραμε, αλλά δεν μπορούν να γνωρίζουν αν θα μεταφερθούν τα αρχεία για εργασία στο σπίτι μέσω φορητής μονάδας αποθήκευσης (usb stick) ή αν θα ανεβούν στο σύννεφο (cloud) ή αν θα υπάρχει οποιοσδήποτε διαμοιρασμός αυτών ώστε να υπάρχει ο κίνδυνος να έχουμε εκτεθειμένα τα δεδομένα μας. Τόνισε πως ακόμη και στο στρατό που λειτουργεί με παρόμοιο τρόπο υπάρχει ο κίνδυνος κυβερνοεπίθεσης και να κλαπούν διάφορα δεδομένα.

Στα μελλοντικά σχέδια της ομάδας του Amnesia σύμφωνα με τον κ. Δημακόπουλο είναι ο τελευταίος αλγόριθμος που θα υπάρξει σε επόμενη αναβάθμιση, που είναι μια εύκολη

εφαρμογή του διαφορικού απόρρητου (differential privacy) πάνω στα σύνολα δεδομένων, που ουσιαστικά αυτό το concept δίνει τη μικρότερη απώλεια πληροφορίας (information loss) πάνω στα δεδομένα. Θέλοντας να μας εξηγήσει τι κάνει ουσιαστικά αυτό το concept, μας ανέλυσε πως: *«Βάζει ένα θόρυβο στα δεδομένα και ουσιαστικά παίζει μία μπλόφα με τον επιτιθέμενο. Δηλαδή αν του πεις του επιτιθέμενου ότι έχει γίνει differential privacy, ότι έχει μπει θόρυβος στα δεδομένα, σημαίνει ότι έχουν μπει ακόμη πχ δύο records στα δεδομένα ως άκυρα. Ο επιτιθέμενος δεν μπορεί να γνωρίζει αν τα δεδομένα αυτά τα δύο είναι ψεύτικα ή αν είναι όλα ψεύτικα και επομένως δεν μπορεί να ξεχωρίσει τα αληθινά από τα ψεύτικα ώστε να τα επιτεθεί ή να τα υποκλέψει. Εγώ το έθεσα πολύ απλοϊκά εδώ αλλά είναι πολύ δυνατός μαθηματικός υπολογισμός όταν γίνεται ανωνυμοποίηση ενός dataset με differential privacy. Είναι ίσως το πιο δυνατό και πιο καινοτόμο μοντέλο που υπάρχει αυτή τη στιγμή στην ανωνυμοποίηση».*

Μία ακόμη ερώτηση που απευθύναμε στον κ. Δημακόπουλο ήταν αν μπορεί ένα ανωνυμοποιημένο αρχείο να επανέλθει στην αρχική του μορφή ώστε να μπορεί κάποιος να διαβάσει το αρχικό σύνολο δεδομένων. Η απάντηση του κ. Δημακόπουλου ήταν πως η ανωνυμοποίηση είναι μονόδρομος (one way). Αν πραγματοποιηθεί, η επιστροφή στο αρχικό dataset δεν είναι εφικτή. Γι' αυτό στην ανωνυμοποίηση με τη χρήση της γενίκευσης υπάρχει απόλυτη προστασία πάνω σε αυτό. Γιατί θα μπορούσε μετά να το πάρει ο επιτιθέμενος και γνωρίζοντας με πιο μοντέλο ανωνυμοποίησης (privacy model) έχει ανωνυμοποιηθεί αυτό, να προσπαθήσει να κάνει αντίστροφη διαδικασία. Εδώ δεν κάνουμε κρυπτογραφία όπου κρύβουμε ένα αρχείο με ένα κλειδί και κάποιος άλλος μπορεί να το διαβάσει χρησιμοποιώντας αυτό το κλειδί. Διαχειριζόμαστε πληροφορίες διασφαλίζοντας τα άτομα που έχουν πάρει μέρος σε διάφορες έρευνες ή των εταιριών που τα έχουνε δώσει, προκειμένου να χρησιμοποιηθούν σε μοντέλα μηχανικής μάθησης. Πρώτα συλλέγουμε πληροφορίες, μετά ανωνυμοποιούμε, έπειτα τα δίνουμε στους ερευνητές και μετά τα εισάγουμε στα μοντέλα μηχανικής μάθησης για να βγάλουμε συμπεράσματα.

Ρωτήσαμε τον κ. Δημακόπουλο ποιες οι ενέργειες έχουν πραγματοποιήσει προκειμένου μελλοντικούς χρήστες να εκπαιδευτούν πάνω στο λογισμικό και γιατί έχουν σταματήσει τα διαδικτυακά σεμινάρια (webinar) από το 2021. Ο κ. Δημακόπουλος μας απάντησε πως υπάρχει πλούσιο αρχειακό υλικό ανεβασμένο στο site του Amnesia και λεπτομερειακή καθοδήγηση για το πως ο χρήστης μπορεί να προβεί σε ανωνυμοποίηση των συνόλων δεδομένων του. Τα webinar μείνανε πίσω γιατί ο κ. Τερροβίτης επαναλάμβανε τα ίδια σενάρια. Αν υπάρξει κάποιο διαφορετικό σενάριο τότε θα ανεβάσουμε κάτι καινούριο.

Τέλος ρωτήσαμε τον κ. Δημακόπουλο αν θα παραμείνει το amnesia για ερευνητικούς σκοπούς ή θα προχωρήσει πιο επαγγελματικά και πιο είναι το όραμα της ομάδας για το συγκεκριμένο λογισμικό. Μας απάντησε πως υπάρχει μια σκέψη από τον κ. Τερροβίτη για μια επίσημη βασική έκδοση (standard version) ώστε να μπορέσει αποκτήσει η εφαρμογή χαρακτήρα ιδιωτικού δικαίου και να μπορέσει να βγει επαγγελματικά και επί πληρωμή αλλά μέχρι στιγμής παραμένει σε ερευνητικό επίπεδο. Όλο το εγχείρημα έγινε με την χρηματοδότηση από την Ευρωπαϊκή Ένωση, αλλά υπάρχει η σκέψη για τη δημιουργία μιας startup εταιρίας και να βγει επαγγελματικά. Ως όραμα θα θέλαμε από εδώ και πέρα να ενταχθεί σε μια ιδιωτική εταιρία το amnesia και να πωλείται. Σίγουρα θα θέλαμε να αναβαθμιστεί και να μην έχει προβλήματα καθόλου και να πωλείται σε εταιρίες στην βιομηχανία της αγοράς (industry) και να διατίθεται και στο δημόσιο τομέα, δηλαδή να υπάρξει μια μεγαλύτερη κινητικότητα στην αγορά. Να ενταχθεί σε μια εταιρία ιδιωτικού δικαίου και μέσω προώθησης (marketing) να μπορεί να ενταχθεί στο δημόσιο και στον ιδιωτικό τομέα. Κυρίως στην Ελλάδα που δεν γνωρίζουν καν το εγχείρημα της ανωνυμοποίησης, αλλά να το προωθήσουμε και στο εξωτερικό ακόμη περισσότερο.

ΚΕΦΑΛΑΙΟ 6

6.1 Η εφαρμογή των τεχνικών ανωνυμοποίησης σε δημόσιους οργανισμούς και ιδιωτικούς φορείς: δυνατότητες και προοπτικές χρήσης, στάσεις και απόψεις

Σε αυτό το κεφάλαιο ανήκουν τα αποτελέσματα της ποιοτικής έρευνας που πραγματοποιήσαμε σε ένα μικρό δείγμα στελεχών δημοσίων οργανισμών και ιδιωτικών φορέων, που συλλέχθηκε με την μέθοδο της σκόπιμης δειγματοληψίας. Ο χώρος που πραγματοποιήσαμε την έρευνα, είναι ο χώρος εργασίας του κάθε συνεντευξιαζόμενου. Αυτή η επιλογή έγινε με γνώμονα την επιθυμία του ερευνητή να πραγματοποιηθούν οι συνεντεύξεις σε περιβάλλοντα οικεία και φιλικά προς τους ερωτώμενους. Η διάρκεια της κάθε συνέντευξης διήρκησε περίπου μία (1) ώρα.

Η συλλογή των δεδομένων έγινε με δομημένη συνέντευξη δια ζώσης, με συγκεκριμένες και κατευθυνόμενες ερωτήσεις και απαντήσεις πάνω σε συγκεκριμένα θέματα. Οι δυσκολίες

που αντιμετωπίσαμε ήταν η έλλειψη χρόνου, διότι απευθυνθήκαμε σε υψηλόβαθμα στελέχη του δημόσιου και του ιδιωτικού τομέα με περιορισμένο ελεύθερο πρόγραμμα καθώς και η άρνηση για συνέντευξη με αντικείμενο το ΓΚΠΔ και τα ευαίσθητα προσωπικά δεδομένα.

Το δείγμα που προσπαθήσαμε να συλλέξουμε είναι από υψηλόβαθμα στελέχη στον δημόσιο και ιδιωτικό τομέα, εξαιτίας της δυνατότητάς τους λόγω της θέσης τους, να ασκήσουν διοίκηση και επιρροή στο υφιστάμενο ανθρώπινο δυναμικό, με γνώμονα τη διαχείριση των δεδομένων των πολιτών, χωρίς να εστιάζουμε συγκεκριμένους εργασιακούς τομείς. Από τον αρχικό αριθμό προσκεκλημένων προς συνέντευξη, ανταποκρίθηκαν πέντε διευθυντές (τρεις διευθυντές δημοσίων οργανισμών, δύο ιδιωτικών οργανισμών) και τέσσερις προϊστάμενοι τμημάτων δημοσίου τομέα. Προσδιοριστικός αρνητικός συντελεστής ήταν η επιφύλαξη ή η άγνοια πάνω σε θέματα και κανόνες προστασίας ευαίσθητων προσωπικών δεδομένων.

Η συλλογή των απαντήσεων πραγματοποιήθηκε με απευθείας καταγραφή σε ηλεκτρονικό έγγραφο που περιείχε τις ερωτήσεις. Κατά τη διάρκεια της κάθε συνέντευξης πραγματοποιήσαμε παρουσίαση του λογισμικού ανωνυμοποίησης με αντίστοιχη επίδειξη ενός υποθετικού σεναρίου ανωνυμοποίησης συνόλου δεδομένων. Αφού ολοκληρώθηκε η διαδικασία, έγινε εκ νέου προβολή και ανάγνωση από τους συνεντευξιαζόμενους των απαντήσεων και στην συνέχεια η υπογραφή του εντύπου συναίνεσής τους .

6.2 Ερευνητικοί στόχοι

Η βασική ιδέα πίσω από την διεξαγωγή των συνεντεύξεων ήταν να παρουσιάσουμε την ανάμειξη των ευαίσθητων προσωπικών δεδομένων γενικά, σε δημόσιους οργανισμούς αλλά και σε ιδιωτικούς φορείς, κατά την παραγωγή υπηρεσιών προς τους πολίτες – πελάτες. Για το λόγο αυτό δεν εστιάσαμε σε κάποιο συγκεκριμένο τομέα αλλά επιλέξαμε σφαιρικά διάφορες εκφάνσεις διοίκησης. Ταυτόχρονα όμως προσπαθήσαμε να καταδείξουμε τόσο τη γνώση και εφαρμογή του πλαισίου του ΓΚΠΔ όσο και την τεκμηρίωση, τα τυχόν προβλήματα και πως αντιμετωπίζονται στην πράξη.

Οι ερωτήσεις επιλέχθηκαν σύμφωνα με το προαναφερθέν σκεπτικό, όμως έπειτα από τις αρνητικές απαντήσεις που δεχτήκαμε στις προσκλήσεις μας για συνέντευξη, κατά την ανακοίνωση του θέματος που ερευνούμε, γενικοποιήσαμε τις ερωτήσεις όσο λιγότερο γινόταν και έτσι προσπαθήσαμε:

- Να ερευνήσουμε το είδος των δεδομένων που διαχειρίζονται
- Αν χρήζουν ανωνυμοποίησης
- Αν εφαρμόζεται κάποιο είδος ανωνυμοποίησης
- Αν έχουν υπάρξει προβλήματα στην υπηρεσία με την τήρηση και χρήση προσωπικών δεδομένων και πως αντιμετωπίζονται κατά την εφαρμογή του ΓΚΠΔ
- Δεξιότητες και προσαρμογή του προσωπικού των οργανισμών και των φορέων σε τεχνικές ανωνυμοποίησης με χρήση τεχνολογιών πληροφορικής

Κύριο μέλημα του ερευνητή ήταν να δημιουργήσει ένα όσο το δυνατόν πιο άνετο και φιλικό περιβάλλον ερωτήσεων ώστε να εγκαταλειφθεί η όποια επιφύλαξη και οι συνομιλητές να αναφερθούν σε όσο το δυνατόν με περισσότερες λεπτομέρειες στο τι εφαρμόζεται στην πραγματικότητα καθώς και για τα τυχόν προβλήματα που αντιμετωπίζουν καθημερινά.

1. Ποιες οι ανάγκες σε σύνολα δεδομένων στην υπηρεσίας σας;

Όλοι ανεξαιρέτως απάντησαν πως διαχειρίζονται δεδομένα μεταξύ των γραφείων που συνεργάζονται ή απαντώντας σε αιτήματα από τις κεντρικότερες διοικήσεις και υπουργεία στα οποία ανήκουν. Οι ανάγκες σε σύνολα δεδομένων έχουν να κάνουν με επεξεργασία στοιχείων φυσικών προσώπων που αφορούν τα ίδια τα υποκείμενα ή αντικείμενα που έχουν στην κατοχή τους όπως πχ κάποιο ακίνητο, σύμφωνα με την ΕΔ Διευθύντρια Διεύθυνσης δασών α. Ο ΓΖ Υποδιευθυντής στην Α' Βάθμια εκπαίδευση, σύνολα δεδομένων μαθητών που προάγονται σε επόμενες τάξεις, ενώ ο ΔΜ Διευθυντής τραπέζης, λίστες πελατών για επεξεργασία και εξιδεικευμένες τραπεζικές υπηρεσίες. Επίσης η ΜΔ Διευθύντρια ιδιωτικού ασφαλιστικού φορέα, θεώρησε αυτονόητη τη διαδικασία συλλογής δεδομένων ηλεκτρονικά καθώς έτσι η εταιρία «έχει καλύτερη εικόνα» των πελατών της. Η ΕΚ στέλεχος στο τμήμα ιθαγένειας του υπουργείου εσωτερικών μας ενημέρωσε πως τα στοιχεία των πολιτών συγκεντρώνονται σε φυσικό φάκελο αλλά και ψηφιακά σε ειδικό λογισμικό του υπουργείου, ενώ τέλος, σύμφωνα με την ΜΣ Προϊσταμένη δημόσιου νοσοκομείου, καταγράφονται όλα τα προσωπικά δεδομένα των ασθενών για την καλύτερη εξυπηρέτηση τους αλλά και για την καλύτερη διαχείριση των πόρων του οργανισμού.

2. Ποιες οι ανάγκες ανωνυμοποίησης σε σύνολα δεδομένων στην υπηρεσίας σας;

Σε αυτήν την ερώτηση οι απαντήσεις ποικίλουν και είναι λογικό διότι προέρχονται από διαφορετικούς εργασιακούς τομείς, όμως έχουν ένα κοινό χαρακτηριστικό. Όλες οι υπηρεσίες

χρησιμοποιούν λογισμικό εσωτερικής λειτουργίας με το οποίο διαμοιράζονται τις πληροφορίες και τα δεδομένα από την καθημερινή εργασία τους.

Σύμφωνα με τον ΔΑ Διευθυντή Δημόσιας Οικονομικής Υπηρεσίας (ΔΟΥ), τα δεδομένα υπόκεινται σε κεντρικό έλεγχο από την ΑΑΔΕ σε επίπεδο ασφαλείας,. Η ΕΚ στέλεχος στο τμήμα ιθαγένειας του υπουργείου εσωτερικών, μας έδειξε το λογισμικό το οποίο χρησιμοποιεί το υπουργείο και καταχωρείται όλος ο φυσικός φάκελος σε ψηφιακή μορφή. Η ΜΔ Διευθύντρια ιδιωτικού ασφαλιστικού φορέα, διευκρίνισε πως όλα τα στοιχεία των πελατών του συστήματός τους, καταχωρούνται από συγκεκριμένους υπολογιστές, οι οποίοι είναι κλειδωμένοι σε αντίστοιχους χρήστες με κωδικούς πιστοποιημένους από τα κεντρικά και οτιδήποτε καταχωρείται πραγματοποιείται σε ένα ασφαλές και κλειστό περιβάλλον. Η ΜΣ Προϊσταμένη δημόσιου νοσοκομείου, μας απάντησε πως στα δεδομένα των ασθενών παρέχεται πρόσβαση επιπέδων, δηλαδή βασική πρόσβαση σε διοικητικούς υπαλλήλους έως και πλήρη πρόσβαση από τους γιατρούς στα στοιχεία των ασθενών, όλα μέσω της εφαρμογής που χορηγείται από την υπηρεσία. Ο ΓΖ Υποδιευθυντής Α' Βάθμιας εκπαίδευσης, εκτός από την καταχώρηση στην σχολική πλατφόρμα (my school), επεξεργάζεται λίστα με όλα τα στοιχεία μαθητών (τύπου excel) προς Α' Βάθμια, με υγειονομικά στοιχεία (covid19), μαθητές ολοήμερου κλπ. Τέλος, η ΕΔ Διευθύντρια Διεύθυνσης δασών δήλωσε πως τα δεδομένα, συνήθως τίτλοι ακινήτων, καταχωρούνται στο αντίστοιχο λογισμικό του κτηματολογίου και αποστέλλονται στις αρμόδιες υπηρεσίες μέσω αυτού, ομοίως ο ΚΜ Προϊστάμενος ανεξάρτητης αρχής, απάντησε πως τα όλα τα δεδομένα που διαχειρίζονται καταχωρούνται στην ειδική πλατφόρμα του υπουργείου και συνήθως πρόκειται για στοιχεία που προέρχονται από τυπικούς ελέγχους ή καταγγελίες. Ομοίως και ο ΔΜ Διευθυντής τραπέζης, δήλωσε πως τα στοιχεία των πελατών καταχωρούνται σε δικό τους τραπεζικό πληροφοριακό σύστημα.

3. Ποια τα προβλήματα που δημιουργεί η αρχή προστασίας προσωπικών δεδομένων στην λειτουργία των υπηρεσιακών διαδικασιών ;

Η ΣΜ Προϊσταμένη πληροφορικής ΔΟΥ και ο ΔΑ Διευθυντής ΔΟΥ, επεσήμαναν πως αναφορικά με την εφαρμογή της νομοθεσίας για την διασφάλιση των προσωπικών δεδομένων στο πλαίσιο του GDPR, υπάρχουν συνεργασίες με συμβουλευτικές εταιρίες οι οποίες παρέχουν εξειδικευμένους υπεύθυνους προστασίας δεδομένων (data protection officer - dpo), που κατευθύνουν τον οργανισμό αναφορικά με την εφαρμογή του κανονιστικού πλαισίου. Η ΜΔ Διευθύντρια ιδιωτικού ασφαλιστικού φορέα, η οποία έδειξε και τον μεγαλύτερο προβληματισμό, επεσήμανε πως παρόλο που η εταιρία δίνει μεγάλη σημασία στους κανόνες

που θέτει η αρχή προστασίας προσωπικών δεδομένων, με συνεχή σεμινάρια επιμόρφωσης, υπάρχουν φορές που δυσκολεύονται να εργαστούν στο πλαίσιο εφαρμογής του ΓΚΠΔ γιατί είναι πολύ αυστηρό και χωρίς δυνατότητες ευελιξίας. Ο ΓΖ Υποδιευθυντής Α' Βάθμιας εκπαίδευσης, δήλωσε πως το πρόβλημα που αντιμετωπίζουνε είναι με την προστασία των δεδομένων όταν αποστέλλουνε μηνύματα ηλεκτρονικού ταχυδρομείου. Ειδικά όταν πρόκειται για αποστολή μηνυμάτων σε γονείς όπου δεν θέλουμε να γνωστοποιήσουμε διευθύνσεις ηλεκτρονικών ταχυδρομείων ή στοιχεία άλλων γονέων ή μαθητών. Μια άλλη οπτική μας ανέλυσε ο ΔΜ Διευθυντής τραπέζης, καθώς δεν αντιμετωπίζει κάποιο πρόβλημα με την αρχή επί της ουσίας, αλλά δήλωσε πως έχουν να αντιμετωπίσουν λειτουργικά προβλήματα από την εφαρμογή του ΓΚΠΔ, δηλαδή κόστος χρηματικό και λειτουργικό, εξαιτίας των μηχανισμών που απαιτούνται για την εφαρμογή και την τήρηση όλων των αυστηρών κανόνων που επιτάσσει ο ΓΚΠΔ.

Όλοι οι υπόλοιποι συνεντευξιαζόμενοι δήλωσαν πως τα δεδομένα τους δεν υπόκεινται σε έλεγχο από την Αρχή προστασίας προσωπικών δεδομένων. Πιο συγκεκριμένα, ο ΚΜ μας είπε πως δεν υπάρχει κάποιο πρόβλημα αναφορικά με την αρχή προστασίας προσωπικών δεδομένων, διότι εμείς καταχωρούμε τα στοιχεία στην ηλεκτρονική πλατφόρμα του υπουργείου. Το ίδιο μας απάντησαν και οι ΕΚ στέλεχος του τμήματος ιθαγένειας στο υπουργείο εσωτερικών, ΕΔ Διευθύντρια Διεύθυνσης δασών και ΜΣ Προϊσταμένη δημόσιου νοσοκομείου, καθώς χρησιμοποιείται η αντίστοιχη ηλεκτρονική πλατφόρμα της υπηρεσίας τους για την καταχώρηση και μετέπειτα επεξεργασία των προσωπικών δεδομένων.

4. Ποια προβλήματα θα μπορούσε να λύσει η χρήση λογισμικού / τεχνικών ανωνυμοποίησης

Σε αυτήν την ερώτηση είχαμε την μεγαλύτερη σύμπνοια απόψεων από τους περισσότερους συνεντευξιαζόμενους. Με ποσοστό της τάξεως του 90%, οι περισσότεροι θεωρούν πως η εφαρμογή τεχνικών ανωνυμοποίησης με εύκολο και αυτοματοποιημένο τρόπο, θα πρόσδιδε αίσθημα ασφάλειας αναφορικά με την διαχείριση των δεδομένων σε καθημερινή βάση και την επικοινωνία με τις άλλες υπηρεσίες. Επίσης θα έδινε τη δυνατότητα επεξεργασίας των πρωτογενών δεδομένων και τη δυνατότητα της τοπικής ή και περιφερειακής χρησιμοποίησής τους για βελτίωση των παρεχόμενων υπηρεσιών.

Πρόσθετα η ΣΜ Προϊσταμένη πληροφορικής ΔΟΥ και η ΜΔ Διευθύντρια ιδιωτικού ασφαλιστικού φορέα, πρότειναν πως ιδανικά θα μπορούσε ένας υπάλληλος του κάθε οργανισμού ειδικότερα κλάδου πληροφορικής, να πιστοποιηθεί ως υπεύθυνος προστασίας των

δεδομένων (dpo), ούτως ώστε γνωρίζοντας τα δυνατά και αδύνατα σημεία του εκάστοτε συστήματος, να μπορεί να έχει την ευθύνη, τον έλεγχο και τη διαχείριση όλων των διαδικασιών που αφορούν τα προσωπικά δεδομένα και τον GDPR.

5. **Υπάρχει δεκτικότητα – διάθεση από τη διοίκηση του φορέα για εκπαίδευση των υπαλλήλων σε τεχνικές ανωνυμοποίησης δεδομένων;**

Εδώ έχουμε μία καθολικότητα θετικών απόψεων αναφορικά με τη δεκτικότητα και τη διάθεση της διοίκησης για την εκπαίδευση των υπαλλήλων. Σύμφωνα με τη ΜΔ Διευθύντρια ιδιωτικού ασφαλιστικού φορέα, θεωρεί πως καμία εταιρία του ιδιωτικού τομέα δεν είναι αρνητική σε νέες τεχνολογίες και εκπαίδευση του προσωπικού, ειδικά αν αυτό συνδράμει στη βελτίωση της παραγωγικότητας. Η ΕΚ στέλεχος του τμήματος ιθαγένειας του υπουργείου εσωτερικών, δήλωσε πως σίγουρα υπάρχει δεκτικότητα, τουλάχιστον από τους νεότερους υπαλλήλους, καθώς θα ήταν θετικό το να μπορεί ο υπάλληλος να διεκπεραιώσει τις υποθέσεις των πολιτών χρησιμοποιώντας μόνο την ψηφιακή πλατφόρμα αφού με αυτή επικοινωνεί με το υπουργείο. Η ΣΜ προϊσταμένη πληροφορικής ΔΟΥ, επεσήμανε πως ιδανικότεροι υπάλληλοι για αυτήν την δουλειά είναι υπάλληλοι των οργανισμών με γνώσεις του κλάδου της πληροφορικής. Επίσης υπάρχουν ήδη υπάλληλοι που έχουν την δυνατότητα και την επιθυμία να ασχοληθούν με αυτό το κομμάτι της διασφάλισης και προστασίας των προσωπικών δεδομένων. Τέλος η ΜΣ Προϊσταμένη δημόσιου νοσοκομείου, δήλωσε πως ως δημόσιοι υπάλληλοι δεν έχουμε την δυνατότητα να αρνηθούμε σε εντολές που έρχονται από ανώτερες αρχές, ούτε μπορούμε να αρνηθούμε τη χρήση λογισμικών προγραμμάτων. Η ανάπτυξη της τεχνολογίας ακόμη και σήμερα το 2023 περισσότερο θα πρέπει να αντιμετωπίζεται ως εργαλείο βοήθειας και διευκόλυνσης των εργασιών και όχι σαν «καταναγκαστικό έργο». Τέλος, αναφέρθηκε από το σύνολο των ερωτηθέντων πως απαιτείται περισσότερος αριθμός υπαλλήλων και καλύτερος καταμερισμός εργασιών, καθώς οι αυξανόμενες αρμοδιότητες των υπαλλήλων θα λειτουργήσουν αρνητικά όχι θετικά στο τελικό αποτέλεσμα που επιθυμούμε να επιτύχουμε.

ΚΕΦΑΛΑΙΟ 7

Συμπεράσματα

Σε αυτό το κεφάλαιο θα αναφερθούμε στα συμπεράσματα της παρούσης εργασίας. Θα παραθέσουμε τα ερευνητικά αποτελέσματα με βάση τους αρχικούς στόχους που είχαμε θέσει.

Η ανωνυμοποίηση των δεδομένων είναι μία δύσκολη διαδικασία που απαιτεί γνώση και εξοικείωση με σύνολα δεδομένων, κανόνες γενίκευσης και αλγόριθμους ανωνυμοποίησης. Είναι μία διαδικασία μονόδρομος και μη αναστρέψιμη, η οποία καθιστά τα πρωτογενή σύνολα δεδομένων φυσικών προσώπων σε μορφή τέτοια, που να μην είναι πλέον αναγνωρίσιμη στην αρχική της μορφή, να παραμένει όμως ικανή για περαιτέρω χρήση, για σκοπούς έρευνας και αξιολόγησης.

Η ανωνυμοποίηση των δεδομένων είναι απαραίτητο εργαλείο για τη διαχείριση των δεδομένων των δημοσίων οργανισμών και των ιδιωτικών φορέων. Με την χρήση της ανωνυμοποίησης δίνεται η δυνατότητα σε οργανισμούς και φορείς, να χρησιμοποιήσουν τα δεδομένα τους χωρίς τον κίνδυνο της διαρροής αυτών. Η μετέπειτα χρήση αυτών των ανωνυμοποιημένων δεδομένων, μπορεί να πραγματοποιηθεί με ασφάλεια, από τον ίδιο τον οργανισμό ή τον φορέα, από εξωτερικά ερευνητικά κέντρα όπως πχ έρευνες από πανεπιστήμια, για στατιστικούς ή ιατρικούς σκοπούς κλπ., με σκοπό την περαιτέρω έρευνα – ανάλυση – αξιολόγηση – ανατροφοδότηση των προϊόντων και υπηρεσιών τους.

Η διαδικασία της ανωνυμοποίησης διευκολύνεται σημαντικά με τη χρήση λογισμικών ανωνυμοποίησης. Πρόκειται για εφαρμογές που κυκλοφορούν στην παγκόσμια αγορά σε ελεύθερη ή επί πληρωμή διάθεση, με ίδιες ή παραπλήσιες τεχνικές, με απλούστερο ή πολυπλοκότερο χειρισμό. Ανάμεσα σε αυτά που ερευνήσαμε ήταν τα Amnesia, ARX, Bizdata-X, G9-Anonymizer, Clover-DX, μ-Argus. Επικρατέστερα όλων ήταν τα Amnesia και ARX λόγω της δωρεάν διάθεσής τους, της πληθώρας των επιλογών που παρείχαν στον ερευνητή και της ευκολίας στην όλη διαδικασία της ανωνυμοποίησης των συνόλων δεδομένων. Εμπορικότερα όλων ήταν το λογισμικό ARX καθώς είναι το παλαιότερο λογισμικό, με τη

μεγαλύτερη διάθεση αλγόριθμων ανωνυμοποίησης και τις περισσότερες επιλογές γενίκευσης δεδομένων από τον ανταγωνισμό.

Σε μεγαλύτερο βάθος ερευνήσαμε το λογισμικό ανωνυμοποίησης δεδομένων Amnesia. Επιλέχθηκε γιατί είναι ελληνικής δημιουργίας, είναι ελεύθερο φτιαγμένο στην πλατφόρμα της java και δωρεάν προς όλους, ενώ διαπιστώθηκε πως έχει τον πιο απλό και προσιτό τρόπο χρήσης για αρχάριους χρήστες. Ταυτόχρονα όμως, επειδή έχει δομικά στοιχεία από το σαφώς παλαιότερο και πιο έμπειρο ARX, μπορεί να χρησιμοποιηθεί και από πιο έμπειρους χρήστες σε ανωνυμοποίηση δεδομένων με ποιο σύνθετες επιλογές. Επίσης πρόκειται για το λογισμικό με την λιγότερη επιβάρυνση επεξεργαστικής ισχύος από τον ηλεκτρονικό υπολογιστή που το εκτελεί.

Στα θετικά συμπεράσματα το Amnesia:

α) έχει το απλούστερο και φιλικότερο γραφικό περιβάλλον από οποιοδήποτε άλλο λογισμικό του ανταγωνισμού

β) χρησιμοποιείται διαδικτυακά (on-line) ή με εγκατάσταση στον Η/Υ

γ) αναγνωρίζει αυτόματα, στις περισσότερες περιπτώσεις, το είδος των χαρακτηριστικών των συνόλων δεδομένων που εισάγει ο χρήστης.

δ) παρέχει τη δυνατότητα στο χρήστη να επιλέξει με ποια χαρακτηριστικά επιθυμεί να προχωρήσει στην ανωνυμοποίηση

ε) προσφέρει στο χρήστη αυτοματοποιημένα (αλλά και χειροκίνητα) τη δυνατότητα γενίκευσης – ομαδοποίησης δεδομένων με τη δημιουργία κανόνων ιεράρχησης αναλόγως του χαρακτηριστικού

στ) εφαρμόζει με απλό και κατανοητό τρόπο την k-ανωνυμία και km-ανωνυμία

ζ) μέσω ενός ειδικού «γραφήματος λύσης» με κόμβους να παρουσιάζει το τελικό ανωνυμοποιημένο σύνολο δεδομένων.

Στα αρνητικά συμπεράσματα το Amnesia:

α) δεν επιτεύχθηκε συμβατότητα με την ελληνική γλώσσα κατά τη διεξαγωγή της παρούσας έρευνας

β) περιορίζεται σε αρχεία συνόλων δεδομένων με μορφή .txt ή csv

γ) δεν δίνεται η επιλογή για αποθήκευση του τελικού ανωνυμοποιημένου αρχείου, σε μορφή διαφορετική από αυτή του αρχικού αρχείου εισαγωγής.

δ) είναι σχεδιασμένο σε java, γεγονός που σημαίνει πως αν δεν ενημερώνεται τακτικά, μπορεί να δημιουργήσει ασυμβατότητες με Η/Υ

ε) χρησιμοποιεί μόνο αλγόριθμους k-ανωνυμίας και km-ανωνυμίας σε αντίθεση με πχ το ARX, που παρέχει σχεδόν κάθε γνωστό αλγόριθμο

στ) βρίσκεται ακόμη σε ερευνητικό επίπεδο χωρίς επαγγελματική έκδοση.

Σύμφωνα με τη συνέντευξη του κ Δημακόπουλου, στέλεχος στην ανάπτυξη και την συντήρηση του λογισμικού Amnesia από το 2019, στον συγγραφέα της παρούσας εργασίας, από τη σκοπιά του δημιουργού:

i. Κύριος σκοπός του λογισμικού Amnesia είναι να παρέχει ασφάλεια των ευαίσθητων προσωπικών δεδομένων στα πλαίσια του ΓΚΠΔ σε οργανισμούς και φορείς
ii. Ο βασικότερος ανταγωνιστής είναι το ARX. Γι' αυτό και το Amnesia δημιουργήθηκε με σκοπό να το ανταγωνιστεί και τα έχει καταφέρει στον τομέα της ευχρηστίας του γραφικού του περιβάλλοντος.
iii. Τα περισσότερα λογισμικά κατασκευάζονται την κεντρική και Βόρεια Ευρώπη. Με την δημιουργία ενός ανταγωνιστικού προϊόντος αναβαθμίζεται και η Ελλάδα στο τεχνολογικό κομμάτι της ανωνυμοποίησης των δεδομένων.

Ως συμπεράσματα από τις συνεντεύξεις με στελέχη οργανισμών και φορέων οφείλουμε να αναφέρουμε τα εξής:

i. Υπήρξε άρνηση για συνέντευξη, σε πολλές από τις προσκλήσεις μας, στην ανακοίνωση του θέματος των ευαίσθητων προσωπικών δεδομένων
ii. Στις περιπτώσεις που οι συνεντευξιαζόμενοι δέχτηκαν τη συμμετοχή τους, υπήρξε ιδιαίτερη επιφύλαξη κατά τη διάρκεια των απαντήσεών τους.
iii. Σε αρκετές περιπτώσεις διαπιστώθηκε ελλιπής γνώση του ΓΚΠΔ ή εμπλοκή του ηθικού περί του τι είναι σωστό και τι σύννομο.

iv. Πολλές υπηρεσίες χρησιμοποιούν εσωτερικά λογισμικά με απευθείας επικοινωνία με τα υπουργεία και τις κεντρικές διοικήσεις
v. Μετά από το τέλος της κάθε συνέντευξης, άρχισαν να δείχνουν μεγαλύτερο ενδιαφέρον για την διαδικασία της ανωνυμοποίησης των προσωπικών δεδομένων

Η εμπειρία του ερευνητή από την παρούσα μελέτη καταλήγει με το συμπέρασμα πως η ανωνυμοποίηση συνόλων δεδομένων είναι μία διαδικασία όχι τόσο γνωστή και διαδεδομένη στην ελληνική πραγματικότητα. Είναι μία διαδικασία πολύπλοκη που απαιτεί εξιδεικευμένες γνώσεις στη σωστή εφαρμογή όλων των κανόνων μέχρι του τελικού αποτελέσματος. Με την εξέλιξη της τεχνολογίας και τη δημιουργία εφαρμογών ανωνυμοποίησης, αυτοματοποιούνται οι διαδικασίες με σκοπό την εξυπηρέτηση των χρηστών – υπαλλήλων, αλλά πρωτίστως προάγεται η ασφάλεια των ίδιων των ευαίσθητων προσωπικών δεδομένων και των υποκειμένων τους στα πλαίσια του ΓΚΠΔ.

Ένα ανωνυμοποιημένο σύνολο δεδομένων είναι στη διάθεση του οργανισμού ή του φορέα που το δημιούργησε, να προχωρήσει σε έρευνα και ανάλυση του παραγόμενου προϊόντος ή υπηρεσίας, ώστε να λειτουργήσει ως ανατροφοδότηση (feedback), με στόχο την εξέλιξη και βελτίωση της παραγωγικής διαδικασίας, της εξοικονόμησης πόρων, της καλύτερης διαχείρισης ανθρώπινου δυναμικού και τελικά ο κύκλος αυτός να κλείσει με ένα νέο καλύτερο τελικό προϊόν ή υπηρεσία.

Από την έρευνα που πραγματοποιήσαμε στο ελληνικής δημιουργίας λογισμικό ανωνυμοποίησης Amnesia, διαπιστώσαμε πως αποτελεί μια φιλόδοξη και καινοτόμο προσπάθεια για τα ελληνικά δεδομένα που δεν έχει να ζηλέψει τίποτα από τα αντίστοιχα του εξωτερικού. Ωστόσο, εντοπίστηκαν προβλήματα και ελλείψεις, με την ευχή να ληφθούν υπόψιν και να βελτιωθούν σε μελλοντική έκδοση.

Ως πρόταση της παρούσας εργασίας είναι ότι θα πρέπει να δοθεί ιδιαίτερη σημασία στην ενημέρωση όλων των υπαλλήλων των οργανισμών και των φορέων σχετικά με τον ΓΚΠΔ, με σεμινάρια και επιμορφώσεις, ώστε αυτή η εσωστρέφεια και άρνηση που δεχτήκαμε να μετατραπεί σε εξωστρέφεια και θετικότητα πάντα με γνώμονα την καλύτερη εξυπηρέτηση του πολίτη – πελάτη και την παραγωγή ενός καλύτερου και ανταγωνιστικότερου προϊόντος – υπηρεσίας.

Σύνοψη της παρούσας εργασίας

Όπως δείξαμε, η διαδικασία της ανωνυμοποίησης, μας δίνει τη δυνατότητα να μετατρέψουμε σύνολα δεδομένων που περιέχουν ευαίσθητα προσωπικά δεδομένα κατά το ΓΚΠΔ, σε μορφή τέτοια ούτως ώστε τα δεδομένα που θα προκύψουν, να μπορέσουν να ενταχθούν σε επιστημονικές έρευνες και να εξαχθούν ερευνητικά δεδομένα.

Αντίστοιχα η διαδικασία της ανωνυμοποίησης κατοχυρώνει την προστασία του πολίτη και των δεδομένων του και εντάσσεται στο κανονιστικό πλαίσιο που θέτει ο ΓΚΠΔ.

Αυτή η μελέτη είχε ως πρωταρχικό σκοπό την κατανόηση εννοιών και νομοθεσιών σχετικών με την προστασία των ευαίσθητων προσωπικών δεδομένων στο κανονιστικό πλαίσιο που επιβάλει ο ΓΚΠΔ. Με την τεχνολογία ως αρωγό παρουσιάσαμε υποθετικά σενάρια ανωνυμοποίησης, με σύγχρονες μεθόδους προστασίας των συνόλων δεδομένων, με τη βοήθεια λογισμικών φιλικών προς το χρήστη και με εγγυημένη την ανωνυμοποίηση τους και την περαιτέρω έρευνα.

Η διαδικασία της ανωνυμοποίησης είναι σύνθετη, πολύπλοκη και κατανοητή σε ειδικό και εξειδικευμένο προσωπικό. Στην πράξη, παρουσιάσαμε λογισμικά, εγχώριας και διεθνούς δημιουργίας, με τα πλεονεκτήματα και τα μειονεκτήματά τους, πως μπορούν να προστατεύσουν όλα τα ψηφιακά δεδομένα από τις άπειρες καθημερινές ψηφιακές συναναστροφές των οργανισμών, των επιχειρήσεων και των ατόμων, με τρόπο απλό, κατανοητό και λειτουργικό για τον οποιοδήποτε μέσο χρήστη αλλά ωστόσο, να ενθαρρύνουν την έρευνα, που αυτή με τη σειρά της θα οδηγήσει στην ανατροφοδότηση και τη βελτίωση της παραγωγικής διαδικασίας.

ΒΙΒΛΙΟΓΡΑΦΙΑ:

Amatriain, X., & Basilico, J. (2015). Recommender systems in industry: A netflix case study. *Recommender Systems Handbook*, 385–419.

Amnesia Anonymization Tool - Data anonymization made easy. Retrieved April 3, 2023, from https://amnesia.openaire.eu/

Ampazis, N. (2010). Large scale problem solving with neural networks: the netflix prize case. *Artificial Neural Networks–ICANN 2010: 20th International Conference, Thessaloniki, Greece, September 15-18, 2010, Proceedings, Part III 20*, 429–434.

AOL | American company | Britannica. (n.d.). Retrieved March 31, 2023, from https://www.britannica.com/topic/AOL

ARX - Data Anonymization Tool | A comprehensive software for privacy-preserving microdata publishing. ARX - Data Anonymization Tool Webpage. Retrieved April 3, 2023, from https://arx.deidentifier.org/

Ayala-Rivera, V., McDonagh, P., Cerqueus, T., & Murphy, L. (2014). A systematic comparison and evaluation of k-anonymization algorithms for practitioners. *Transactions on Data Privacy*, *7*(3), 337–370.

Barbaro, M., Zeller, T., & Hansell, S. (2006). A face is exposed for AOL searcher no. 4417749. *New York Times*, *9*(2008), 8.

Bolognini, L., & Bistolfi, C. (2017). Pseudonymization and impacts of Big (personal/anonymous) Data processing in the transition from the Directive 95/46/EC to the new EU General Data Protection Regulation. *Computer Law & Security Review*, *33*(2), 171–181.

Buneman, P. (1997). Semistructured data. *Proceedings of the Sixteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, 117–121.

Chandrakar, P., & Om, H. (2015). RSA based two-factor remote user authentication scheme with user anonymity. *Procedia Computer Science*, *70*, 318–324.

Differential privacy accounting by connecting the dots – Google AI Blog. (n.d.). Retrieved March 31, 2023, from https://ai.googleblog.com/2022/12/differential-privacy-accounting-by.html

Differential Privacy A privacy-preserving system. Last Retrieved March 15, 2023, from ''https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf''.

Differential Privacy | Harvard University Privacy Tools Project. Retrieved March 31, 2023, from https://privacytools.seas.harvard.edu/differential-privacy

Dwork, C., & Rothblum, G. N. (2016). Concentrated differential privacy. *ArXiv Preprint ArXiv:1603.01887*.

EUR-Lex - 32016R0679 - EN - EUR-Lex. Retrieved April 3, 2023, from https://eur-lex.europa.eu/eli/reg/2016/679/oj

Gkountouna, O., Angeli, S., Zigomitros, A., Terrovitis, M., & Vassiliou, Y. (2014). km-Anonymity for continuous data using dynamic hierarchies. *Privacy in Statistical Databases: UNESCO Chair in*

Data Privacy, International Conference, PSD 2014, Ibiza, Spain, September 17-19, 2014. Proceedings, 156–169.

Haber, A. C., Sax, U., Prasser, F., & Consortium, N. (2022). Open tools for quantitative anonymization of tabular phenotype data: literature review. *Briefings in Bioinformatics*, *23*(6), bbac440.

Hillyard, D., & Gauen, M. (2007). Issues around the protection or revelation of personal information. *Knowledge, Technology & Policy*, *20*, 121–124.

How AOL dominated the internet of the '90s and let it slip away. (n.d.). Retrieved March 31, 2023, from https://www.cnbc.com/2019/08/15/how-aol-dominated-the-internet-of-the-90s-and-let-it-slip-away.html

International Classification of Diseases (ICD). (n.d.). Retrieved April 20, 2023, from https://www.who.int/standards/classifications/classification-of-diseases

Julkunen, H., & Ceder Molander, J. (2016). *Password strength and memorability*.

Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, *37*(3), 50–60.

Marques, J. F., & Bernardino, J. (2020). Analysis of Data Anonymization Techniques. *KEOD*, 235–241.

Meurers, T., Bild, R., Do, K.-M., & Prasser, F. (2021). A scalable software solution for anonymizing high-dimensional biomedical data. *GigaScience*, *10*(10), giab068.

Murthy, S., Bakar, A. A., Rahim, F. A., & Ramli, R. (2019). A comparative study of data anonymization techniques. *2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing,(HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*, 306–309.

Narayanan, A., & Shmatikov, V. (2008). Robust De-anonymization of Large Datasets (How to Break Anonymity of the Netflix Prize Dataset). The University of Texas at Austin. *Proceedings of the 29th IEEE Symposium on Security and Privacy, Oakland, CA, USA*, 18–21.

Overview - Microdata - Eurostat. (n.d.). Retrieved May 4, 2023, from https://ec.europa.eu/eurostat/web/microdata

Peter R. Cox. (1976). *Demography*. Cambridge University Press.

Prasser, F., Kohlmayer, F., Lautenschläger, R., & Kuhn, K. A. (2014). Arx-a comprehensive tool for anonymizing biomedical data. *AMIA Annual Symposium Proceedings*, *2014*, 984.

Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., & Maier-Hein, K. (2020). The future of digital health with federated learning. *NPJ Digital Medicine*, *3*(1), 119.

Rodríguez-Barroso, N., Jiménez-López, D., Luzón, M. V., Herrera, F., & Martínez-Cámara, E. (2023). Survey on federated learning threats: Concepts, taxonomy on attacks and defences, experimental study and challenges. *Information Fusion*, *90*, 148–173.

Rodwald, P. (2021). Effectiveness Comparison of Email Addresses Recovery from Gravatars. *Web Engineering: 21st International Conference, ICWE 2021, Biarritz, France, May 18–21, 2021, Proceedings*, 505–508.

Simi, M. S., Nayaki, K. S., & Elayidom, M. S. (2017). An extensive study on data anonymization algorithms based on k-anonymity. *IOP Conference Series: Materials Science and Engineering*, *225*(1), 012279.

Sweeney, L. (2000). Simple demographics often identify people uniquely. *Health (San Francisco)*, *671*(2000), 1–34.

Sweeney, L. (2002). Achieving k-anonymity privacy protection using generalization and suppression. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *10*(05), 571–588.

Teece, D. J. (2018). SWOT Analysis. In M. Augier & D. J. Teece (Eds.), *The Palgrave Encyclopedia of Strategic Management* (pp. 1689–1690). Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-00772-8_285

Terrovitis, M., Mamoulis, N., & Kalnis, P. (2011). Local and global recoding methods for anonymizing set-valued data. *The VLDB Journal*, *20*, 83–106.

Torra, V., & Navarro-Arribas, G. (2016). Big data privacy and anonymization. *Privacy and Identity Management. Facing up to Next Steps: 11th IFIP WG 9.2, 9.5, 9.6/11.7, 11.4, 11.6/SIG 9.2. 2 International Summer School, Karlstad, Sweden, August 21-26, 2016, Revised Selected Papers 11*, 15–26.

United Nations Development Group. (2017). *DATA PRIVACY, ETHICS AND PROTECTION GUIDANCE NOTE ON BIG DATA FOR ACHIEVEMENT OF THE 2030 AGENDA*. Retrieved March 15, 2023 from ''https://unsdg.un.org/sites/default/files/UNDG_BigData_final_web.pdf''

University of North Alambama. *Security Incidents around Us*. Retrieved April 3, 2023, from https://www.una.edu/its/technology-security/known-security-incidents.html

Wolf, D., Henley, A. J., Wolf, D., & Henley, A. J. (2017). Use gravatar to display user's avatars with posts. *Java EE Web Application Primer: Building Bullhorn: A Messaging App with JSP, Servlets, JavaScript, Bootstrap and Oracle*, 123–125.

Wu, X., Zhang, Y., Shi, M., Li, P., Li, R., & Xiong, N. N. (2022). An adaptive federated learning scheme with differential privacy preserving. *Future Generation Computer Systems*, *127*, 362–372.

Yıldırım, M., & Mackie, I. (2019). Encouraging users to improve password security and memorability. *International Journal of Information Security*, *18*, 741–759.

Yin, X., He, J., Guo, Y., Han, D., Li, K.-C., & Castiglione, A. (2020). An efficient two-factor authentication scheme based on the Merkle tree. *Sensors*, *20*(20), 5735.

ΑΠΔΠΧ Γνωμοδότηση 1/2017. (n.d.). *ΑΠΔΠΧ Γνωμοδότηση 1/2017*.

ΑΠΔΠΧ Γνωμοδότηση 4/2017. (n.d.). *ΑΠΔΠΧ Γνωμοδότηση 4/2017*.

Γιανναντίδης, Θ. Ρ. (2022). *Το παράδοξο της ιδιωτικότητας των δεδομένων έναντι της εξατομίκευσης περιεχομένου: Μία ανάλυση μέσω διάθεσης διαμοιρασμού προσωπικών δεδομένων και κλιμάκων εμπειρίας του πελάτη*. https://hephaestus.nup.ac.cy/handle/11728/12176

Γκαδόλος, Ι. (1998). Ασφάλεια Δια-Δικτυακής Διακίνησης Πληροφοριών. *Πτυχιακή Εργασία, ΤΕΙ Πειραιά*. ''platon.teipir.gr/new/ecs/pelab_1/ptixiakes/jgkadol.pdf''

Δημοπούλου, Σ. (2022). *Προστασία προσωπικών δεδομένων από το σχεδιασμό και τεχνολογίες ενίσχυσης της ιδιωτικότητας*. https://doi.org/10.26267/UNIPI_DIONE/1894

Ευρωπαϊκό Κοινοβούλιο (2016). Ευρωπαϊκός Κανονισμός 679/2016 για την προστασία των προσωπικών δεδομένων. ''https://eur-lex.europa.eu/legal-content/EL/TXT/PDF/?uri=CELEX:32016R0679&from=HR''

Ηλεκτρονικό εισιτήριο: Κινδύνους για τα προσωπικά δεδομένα εντοπίζει η ΑΠΔΠΧ. Lawspot. Retrieved April 3, 2023, from https://www.lawspot.gr/nomika-nea/ilektroniko-eisitirio-kindynoys-gia-ta-prosopika-dedomena-entopizei-i-apdph

Καρράς, Σ. Ι., Karras, S. I., Πανοπούλου, Φ.-Β. Π., & Panopoulou, F.-V. P. (2014). *Εργαλείο Ανωνυμοποίησης για Δημοσιεύσεις Δεδομένων*. https://doi.org/10.26240/HEAL.NTUA.4412

Κατσαβριά, Δ. (2018). *Εφαρμογή και επιπτώσεις του γενικού κανονισμού προστασίας δεδομένων (GDPR) στην ελληνική πραγματικότητα (μελέτη περιπτώσεων)*. ''https://dspace.lib.uom.gr/bitstream/2159/22384/4/katsavriaDimitraMsc2018.pdf''

Λιμνιώτης Κ. (2018). *Η ψευδωνυμοποίηση στον Γενικό Κανονισμό Προστασίας Δεδομένων*. https://www.enisa.europa.eu/events/personal-data-security/pseydonymization

Μαρτσούκου Ε.(2018). *ΑΡΧΗ ΠΡΟΣΤΑΣΙΑΣ ΔΕΔΟΜΕΝΩΝ ΠΡΟΣΩΠΙΚΟΥ ΧΑΡΑΚΤΗΡΑ. Ηλεκτρονικό εισιτήριο: Κινδυνεύει η Ιδιωτικότητα των Επιβατών; ''https://www.dpa.gr/sites/default/files/2020-06/MARTSOUKOU_LAMBRINOUDAKIS.PDF''*

Σιουγλέ, Δ., Εεπ, Ε., & Πληροφορικός, Ε. *«Ενίσχυση της ασφάλειας δεδομένων στο ΓΚΠΔ: ψευδωνυμοποίηση-κρυπτογράφηση»*. Διαθέσιμο 04/2023. https://elearning.ekdd.gr/pluginfile.php/26223/mod_folder/content/0/4-%CE%A8%CE%B5%CF%85%CE%B4%CF%89%CE%BD%CF%85%CE%BC%CE%BF%CF%80%CE%BF%CE%AF%CE%B7%CF%83%CE%B7%20%CE%BA%CE%B1%CE%B9%20%CE%BA%CF%81%CF%85%CF%80%CF%84%CE%BF%CE%B3%CF%81%CE%AC%CF%86%CE%B7%CF%83%CE%B7.pdf?forcedownload=1

Στεφανιδάκης, Μ. (2016). *Αλγόριθμοι και Δομές Δεδομένων (ΙΙ) (Γράφοι και Δένδρα)*. Ιόνιο Πανεπιστήμιο - Τμήμα Πληροφορικής. https://docplayer.gr/47177381-Algorithmoi-kai-domes-dedomenon-ii-grafoi-kai-dendra.html

Συνδουκάς, Δ., Τμήμα, Δ., & Επιχειρήσεων, Δ. *ΔΟΜΕΣ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΑΛΓΟΡΙΘΜΟΙ*. Προσπελάστηκε 10 Μαρτίου 2023. ''https://openclass.teiwm.gr/courses/BA-G121/''