

ΕΚΘΕΣΗ

Μεγάλα Δεδομένα (Big Data) στην Υγεία

Εισηγητές: Τάκης Βιδάλης, Βασιλική Μολλάκη



**ΕΘΝΙΚΗ ΕΠΙΤΡΟΠΗ ΒΙΟΗΘΙΚΗΣ**

Νεοφύτου Βάμβα 6, Τ.Κ. 10674, Αθήνα, τηλ. 210- 88.47.700, φαξ 210- 88.47.701

E-mail: [secretariat@bioethics.gr](mailto:secretariat@bioethics.gr), url: [www.bioethics.gr](http://www.bioethics.gr)



## Περιεχόμενα

<b>1. Εισαγωγή.....</b>	<b>4</b>
<b>2. Τα δεδομένα.....</b>	<b>5</b>
2.1. Τα χαρακτηριστικά των μεγάλων δεδομένων .....	5
2.2. Οι πιθανές πηγές δεδομένων για την υγεία .....	5
2.3. Τα οφέλη των μεγάλων δεδομένων στην υγεία .....	7
2.4. Οι πιθανοί κίνδυνοι της τεχνολογίας μεγάλων δεδομένων .....	8
<b>3. Τα ηθικά ζητήματα.....</b>	<b>9</b>
3.1. Το ζήτημα της συναίνεσης .....	9
3.2. Η διασφάλιση της εμπιστευτικότητας των δεδομένων .....	12
<b>4. Η σχετική νομοθεσία.....</b>	<b>12</b>
<b>5. Προτάσεις.....</b>	<b>15</b>
5.1. Πολιτική των φορέων που κάνουν χρήση της τεχνολογίας μεγάλων δεδομένων .....	15
5.2. Μέτρα προστασίας των δεδομένων .....	16
5.3. Συγκατάθεση ύστερα από ενημέρωση .....	16
<b>ΒΙΒΛΙΟΓΡΑΦΙΑ .....</b>	<b>18</b>

## 1. Εισαγωγή

Τα «*big data*» ή αλλιώς «μαζικά» ή «μεγάλα δεδομένα» αποτελούν συλλογή από μεγάλα και περίπλοκα σύνολα δεδομένων τα οποία, όταν αναλυθούν, μπορεί να οδηγήσουν σε συσχετίσεις και συμπεράσματα που θα ήταν ανέφικτα με την ανάλυση μεμονωμένων δεδομένων. Συνεπώς, τα μεγάλα δεδομένα δίνουν τη δυνατότητα να βρεθούν λύσεις σε δύσκολα προβλήματα.

Η διάδοση και η χρήση των μεγάλων δεδομένων στον τομέα της υγείας διαρκώς επεκτείνονται, προκειμένου να συλλεχθούν επιδημιολογικά στοιχεία ή να μελετηθούν σπάνιες ή περίπλοκες ασθένειες. Συνήθεις πηγές πληροφοριών υγείας αποτελούν τα συστήματα ηλεκτρονικού φακέλου, τα συστήματα ηλεκτρονικής συνταγογράφησης, αλλά και οι μεγάλες βάσεις δεδομένων που χρησιμοποιούνται στην έρευνα και στις κλινικές μελέτες. Εντούτοις, πληροφορίες για καθημερινές δραστηριότητες και την υγεία των προσώπων είναι επίσης δυνατόν να αντληθούν από τις αναρτήσεις τους σε μέσα κοινωνικής δικτύωσης και τα ηλεκτρονικά αρχεία αγορών τους.

Μολονότι τα μεγάλα δεδομένα έχουν σημασία σε πολλούς τομείς, η δυναμική τους στη βιοϊατρική, τόσο στην κλινική πράξη όσο και στην έρευνα, είναι σπουδαία. Ωστόσο η χρήση τους εγείρει ηθικά ζητήματα, κυρίως λόγω της απουσίας συγκατάθεσης για τη χρήση των ευαίσθητων πληροφοριών για την υγεία των προσώπων στα οποία ανήκουν τα δεδομένα, λόγω του κινδύνου διαρροής πληροφοριών κατά την αποθήκευση και μεταφορά των δεδομένων αλλά και εξαιτίας ασαφειών ως προς το καθεστώς της ιδιοκτησίας των δεδομένων.

## 2. Τα πραγματικά στοιχεία

### 2.1. Τα χαρακτηριστικά των μεγάλων δεδομένων

Τα μεγάλα δεδομένα χαρακτηρίζονται από τα λεγόμενα «4Vs+1»: **Volume** (όγκο), **Velocity** (ταχύτητα), **Variety** (ποικιλία), **Veracity** (εγκυρότητα) και **Value** (αξία). Χαρακτηρίζονται δηλαδή, από τον μεγάλο όγκο των πληροφοριών (που μπορεί να είναι δομημένες ή μη), τη μεγάλη ταχύτητα με την οποία συλλέγονται και αναλύονται τα δεδομένα, την αυξημένη πολυπλοκότητα και ποικιλομορφία των δεδομένων που φτάνουν σε διαφορετική μορφή (format), την αξιοπιστία τους, αλλά και την αξία που παράγεται από την ανάλυσή τους.

Η ανάλυση των μεγάλων δεδομένων απαιτεί ειδικές στατιστικές τεχνικές και εξειδικευμένους επιστήμονες πληροφορικής, ενώ, ειδικά στην επεξεργασία μεγάλων δεδομένων υγείας, απαραίτητη είναι η συνεισφορά επιστημόνων βιοπληροφορικής.

Η χρήση και η επεξεργασία των μεγάλων δεδομένων στην υγεία πραγματοποιείται στις περισσότερες των περιπτώσεων σε **ανωνυμοποιημένα** ή **ψευδοανωνυμοποιημένα** δεδομένα, ώστε να προστατεύεται η ιδιωτικότητα των μεμονωμένων προσώπων στα οποία αναφέρονται τα στοιχεία, αλλά και να εξυπηρετούνται οι σκοποί της επεξεργασίας. Ανώνυμα είναι τα δεδομένα που δεν επιτρέπουν τον προσδιορισμό ή την ταυτοποίηση των προσώπων στα οποία ανήκουν, ακόμη και αν συνδυαστούν με άλλα στοιχεία από οποιαδήποτε άλλη πηγή. Η ψευδοανωνυμοποίηση των δεδομένων σημαίνει την αντικατάσταση τυχόν χαρακτηριστικών ταυτοποίησης των δεδομένων με ψευδώνυμο ή, με άλλα λόγια, μια τιμή που δεν επιτρέπει την άμεση αναγνώριση του υποκειμένου των δεδομένων.

### 2.2. Οι πιθανές πηγές δεδομένων για την υγεία

Τις τελευταίες δεκαετίες οι εξελίξεις είναι αλματώδεις ως προς τις τεχνικές δυνατότητες για συλλογή, διακίνηση και ανάλυση μεγάλων συνόλων πληροφοριών.

Η ευρεία χρήση του διαδικτύου, η δημιουργία μεγάλων βάσεων δεδομένων για τη βιοιατρική και η ανάπτυξη εφαρμογών κινητών συσκευών έχουν ευνοήσει τη συλλογή δεδομένων. Ενδεικτικές πηγές από όπου μπορούν να αντληθούν πληροφορίες και να αναλυθούν αναφέρονται παρακάτω.

**α) Βάσεις επιστημονικών δεδομένων και μητρώα μελετών** που περιλαμβάνουν κλινικά και εργαστηριακά δεδομένα και, είτε είναι ελεύθερα προσβάσιμες στο διαδίκτυο είτε είναι κλειστές και τα δεδομένα αποθηκεύονται τοπικά σε εργαστήρια. Ιδιαίτερα ο όγκος των παραγόμενων γενετικών δεδομένων αναμένεται να αυξηθεί κατακόρυφα στο μέλλον στις βάσεις δεδομένων, λόγω της ανάπτυξης νέων τεχνολογιών αλληλούχισης και της μείωσης του κόστους της τελευταίας.<sup>1</sup> Σε αυτές τις βάσεις περιλαμβάνονται, επίσης, και τα μητρώα κλινικών μελετών ή δοκιμών στον άνθρωπο, που έχουν διενεργηθεί στο παρελθόν ή διενεργούνται στο παρόν. Τέλος, στις βάσεις αυτές συγκαταλέγεται και η διεθνής επιστημονική βιβλιογραφία.

**β) Κλινικά και διοικητικά αρχεία**, και κυρίως ο ηλεκτρονικός φάκελος υγείας, που περιλαμβάνει επιδημιολογικά στοιχεία, το ιστορικό ασθενειών αλλά και πιθανά θεραπευτικά σχήματα και την ανταπόκριση των ασθενών.<sup>2</sup> Το ηλεκτρονικό σύστημα συνταγογράφησης και τα αρχεία των φορέων ασφάλισης αποτελούν, επίσης, σημαντική πηγή δεδομένων υγείας.

**γ) Κινητές συσκευές και εφαρμογές για την υγεία**, όπως τα έξυπνα τηλέφωνα, τα έξυπνα ρολόγια αλλά και τα έξυπνα βραχιόλια υγείας (smart health bands) που αποτελούν αναδυόμενες τεχνολογίες παρακολούθησης της προσωπικής υγείας.<sup>3,4</sup>

**δ) Μη επιστημονικές πηγές**, όπως είναι οι καθημερινές ηλεκτρονικές αγορές ενός προσώπου, οι καθημερινές αναζητήσεις του στο διαδίκτυο αλλά και οι αναρτήσεις του σε μέσα κοινωνικής δικτύωσης. Χαρακτηριστικό παράδειγμα αποτελεί η ιστορία της Google Flu Trends. Το 2008 οι ερευνητές της Google συνειδητοποίησαν ότι υπήρχε αυξανόμενος αριθμός αναζητήσεων για τη γρίπη, και λαμβάνοντας τη διαδικτυακή διεύθυνση (IP) κάθε χρήστη μπορούσαν να προβλέψουν πιθανές επιδημίες της γρίπης σε διάφορες περιοχές των ΗΠΑ.<sup>5</sup> Σε

ορισμένες χρονικές στιγμές οι προβλέψεις της Google Flu Trends συνέπεσαν με τις προβλέψεις του αμερικανικού Κέντρου Ελέγχου και Πρόληψης Νοσημάτων (Centers for Disease Control and Prevention, CDC).<sup>6</sup> Γνωστά παραδείγματα χρήσης μεγάλων δεδομένων για την πρόβλεψη τάσεων στη δημόσια υγεία αποτελούν το Flu Trends,<sup>5</sup> το Twitter Influenza Surveillance<sup>7</sup> και το Facebook Emotional Contagion study.<sup>8</sup>

Τα παραπάνω προγράμματα δέχτηκαν μεγάλη κριτική κυρίως λόγω της απουσίας συγκατάθεσης των προσώπων, των οποίων τα δεδομένα χρησιμοποιήθηκαν για τις έρευνες. Αφενός μεν τα παραδείγματα αυτά αναδεικνύουν τις δυνατότητες των μεγάλων δεδομένων στην υγεία, αφετέρου δε εγείρονται ζητήματα για την ηθική της χρήσης τους, την προστασία της ιδιωτικής ζωής του κοινού, το απόρρητο των δεδομένων υγείας και την αυτονομία όσων συμμετέχουν στην έρευνα. Αν και τα δεδομένα αυτά αναρτώνται δημοσίως από τους χρήστες των μέσων κοινωνικής δικτύωσης, ωστόσο συλλέγονται και αναλύονται χωρίς τη συγκατάθεση τους. Το κοινό ενημερώνεται ολοένα και περισσότερο για τις πιθανότητες χρήσης των δεδομένων τους για άλλους σκοπούς, αλλά συχνά αγνοεί το δυναμικό των πληροφοριών αυτών.

### **2.3. Τα οφέλη των μεγάλων δεδομένων στην υγεία**

Στον τομέα της υγείας, τα μεγάλα δεδομένα είναι δυνατόν να βελτιώσουν τη λήψη αποφάσεων, να βοηθήσουν στην πρόληψη ασθενειών και εξάπλωσης νόσων αλλά και να μειώσουν το κόστος της ιατροφαρμακευτικής περίθαλψης, εξαιτίας των χαρακτηριστικών τους (4Vs+1).

Η δημιουργία μεγάλων δεδομένων παρέχει τη δυνατότητα σε ιατρούς και ερευνητές να συλλέξουν ένα μεγάλο όγκο πληροφοριών και να τις μοιραστούν ευρέως με άλλα μέλη της επιστημονικής κοινότητας. Η ανάλυση μεγάλων δεδομένων είναι δυνατόν να βοηθήσει να συσχετιστούν διάφορες παράμετροι με ασθένειες, όπως π.χ. να εντοπιστούν νέοι βιοδείκτες που θα προβλέψουν -ως ένα βαθμό- εάν ένα άτομο νοσήσει και πότε. Αυτό έχει ως αποτέλεσμα να καθίσταται δυνατή η μελέτη σύνθετων και σπάνιων ασθενειών σε βάθος και σε σύντομο χρόνο,

οι οποίες μέχρι τώρα ήταν δύσκολο να μελετηθούν με την ανάλυση μικρών συνόλων δεδομένων.

Ιδιαίτερα τα δεδομένα του ηλεκτρονικού φακέλου υγείας και η ανάλυσή τους παρέχουν τη δυνατότητα να αναγνωρισθούν νέοι παράγοντες κινδύνου, να οριστούν επιδημιολογικά στοιχεία, να υπολογιστεί το προσδόκιμο επιβίωσης για διάφορες ασθένειες, να καταγραφούν και να αποφευχθούν ανεπιθύμητες ενέργειες, ακόμη και να αναγνωριστούν και να μετριαστούν τα ιατρικά σφάλματα. Γενικότερα, ο ηλεκτρονικός φάκελος υγείας, και η ηλεκτρονική συνταγογράφηση ευνοούν την καταγραφή, αποστολή και ανταλλαγή δεδομένων μεταξύ επαγγελματιών υγείας για κλινικούς, επιδημιολογικούς, διοικητικούς και ερευνητικούς σκοπούς. Συνεπώς, η ανάλυση των μεγάλων δεδομένων που προκύπτουν από τα συστήματα ηλεκτρονικού φακέλου και ηλεκτρονικής συνταγογράφησης μπορούν να βοηθήσουν στη διαμόρφωση των πολιτικών υγείας που θα πρέπει να εφαρμοστούν.<sup>2</sup>

Στη βιβλιογραφία, ήδη έχουν περιγραφεί ορισμένα παραδείγματα ανάλυσης μεγάλων δεδομένων υγείας, με πολλά υποσχόμενα αποτελέσματα για το όφελος των μεγάλων δεδομένων. Οι μελέτες αυτές περιλάμβαναν κυρίως, τη συλλογή και ανάλυση δεδομένων από ηλεκτρονικούς φακέλους υγείας.<sup>9,10,11,12,13</sup>

Επιπλέον, η ανάλυση μεγάλων δεδομένων παρέχει τη δυνατότητα καλύτερης παρακολούθησης της δημόσιας υγείας, συνδυάζοντας γεωγραφικές και κοινωνικές μεταβλητές με δεδομένα υγείας.<sup>14</sup>

#### **2.4. Οι πιθανοί κίνδυνοι της τεχνολογίας μεγάλων δεδομένων**

Τα μεγάλα δεδομένα ενέχουν κινδύνους που αφορούν ηθικά ζητήματα, όπως αναλύονται παρακάτω, ορισμένα οργανωτικά ζητήματα αλλά και τεχνικές ή μεθοδολογικές δυσκολίες στην επεξεργασία τους.<sup>15</sup> Τα οργανωτικά και τεχνικά ζητήματα υπερβαίνουν τον σκοπό της παρούσας μελέτης. Ωστόσο είναι κρίσιμο να σημειωθεί ο κίνδυνος από-ταυτοποίησης (de-identification) των προσώπων, δηλαδή η πιθανότητα τα δεδομένα να συνδυαστούν με άλλα δεδομένα που τυχόν υπάρχουν



ελεύθερα στο διαδίκτυο ή σε άλλου είδους αρχεία, καταλήγοντας στην ταυτοποίηση των ατόμων. Ο κίνδυνος από-ταυτοποίησης είναι γνωστός σε όλους τους επιστήμονες που ασχολούνται ειδικά με τα μεγάλα δεδομένα, και αυτός ακριβώς είναι ο λόγος για τον οποίο λαμβάνονται συγκεκριμένα μέτρα για να τον εμποδίσουν. Η βιβλιογραφία περιλαμβάνει μελέτες που καταδεικνύουν ότι η ανωνυμοποίηση ή η ψευδοανωνυμοποίηση δεδομένων δεν επαρκεί για να αποτρέψει την από-ταυτοποίηση.<sup>16,17,18,19,20</sup>

### **3. Τα ηθικά ζητήματα**

Δεν υπάρχει αμφιβολία ότι οι βάσεις μεγάλων δεδομένων στον τομέα της Υγείας μπορούν να συντελέσουν καθοριστικά τόσο στην ανάπτυξη νέων φαρμάκων και ιατροτεχνολογικών προϊόντων ή μεθόδων προς όφελος της θεραπείας και της πρόληψης, όσο και στον εξορθολογισμό των οικονομικών και ευρύτερα της δημόσιας πολιτικής για την Υγεία. Η ωφέλεια αυτή, ωστόσο, φαίνεται να μετριάζεται από κινδύνους που προκύπτουν στο επίπεδο της αυτονομίας και των εγγυήσεων της ιδιωτικότητας των υποκειμένων αυτών των ευαίσθητων πληροφοριών. Η εναρμόνιση, επομένως, της αξίας της προστασίας της υγείας με τις δύο τελευταίες αξίες, είναι αναγκαία.

Η δυνατότητα συγκέντρωσης μεγάλου όγκου προσωπικών δεδομένων υγείας, είτε για την εξυπηρέτηση κλινικών εφαρμογών ή πολιτικών δημόσιας υγείας είτε για χρήση σε ερευνητικούς σκοπούς, προκαλεί έναν διπλό προβληματισμό από την άποψη της βιοηθικής α) ως προς τη διασφάλιση της συναίνεσης των υποκειμένων των δεδομένων, και β) ως προς την τήρηση της εμπιστευτικότητας των τελευταίων.

#### **3.1. Το ζήτημα της συναίνεσης**

Οι συλλογές μεγάλων δεδομένων δημιουργούνται κυρίως με μεθόδους διασύνδεσης ήδη υπαρχόντων αρχείων σε μονάδες που διατηρούν πρωτογενείς

συλλογές δεδομένων υγείας (νοσοκομεία, κλινικές, κέντρα υγείας, διαγνωστικά κέντρα, εργαστήρια αναλύσεων, ασφαλιστικούς οργανισμούς κ.λπ.). Οι πρωτογενείς αυτές συλλογές βασίζονται στη ρητή και ειδική συναίνεση των προσώπων από τα οποία προέρχονται τα δεδομένα. Τα πρόσωπα αυτά έχουν ενημερωθεί εκ των προτέρων για τον σκοπό της συλλογής και διατήρησης των δεδομένων: η συναίνεσή τους, επομένως, καλύπτει μόνον τον συγκεκριμένο σκοπό και όχι άλλον.

Αν, λοιπόν, πρόκειται να υπάρξει διασύνδεση αρχείων των πρωτογενών συλλογών, προκειμένου να εξυπηρετηθούν ευρύτεροι σκοποί (πέρα από τον συγκεκριμένο, στον οποίον αφορά η κάθε ειδική συναίνεση), η διασφάλιση της αυτονομίας του προσώπου προϋποθέτει κατ' αρχήν νέα συναίνεση προσαρμοσμένη πλέον στους σκοπούς αυτής της διασύνδεσης αρχείων. Το πρακτικό πρόβλημα που τίθεται, εδώ, είναι ότι, κατά κανόνα, η διαδικασία «επιστροφής» στο κάθε υποκείμενο πρωτογενών δεδομένων είναι δύσκολη, είτε λόγω του μεγάλου όγκου που επιδιώκεται να συγκεντρωθεί είτε λόγω αδυναμίας εντοπισμού προσώπων που είχαν δώσει δεδομένα στο παρελθόν. Αν δεν αντιμετωπισθεί αυτό το πρόβλημα, η δημιουργία βάσεων μεγάλων δεδομένων καθίσταται τελικά ανέφικτη.

Εκείνο που ερωτάται, ειδικότερα, είναι αν θα μπορούσε να νοηθεί αποτελεσματική διασφάλιση της αυτονομίας του προσώπου με εναλλακτικούς τρόπους, πέρα από τον καθιερωμένο της απαίτησης ειδικής και προηγούμενης συναίνεσης, ώστε να διευκολύνεται η ασφαλής διασύνδεση αρχείων προς όφελος ευρύτερων σκοπών στην Υγεία και την Έρευνα.

Μια εναλλακτική επιλογή είναι αυτή της δημιουργίας μεγάλων συλλογών, με βάση τη λεγόμενη «εικαζόμενη συναίνεση» των δοτών. Η επιλογή αυτή υιοθετήθηκε στη γνωστή περίπτωση της ίδρυσης βιοτράπεζας της Ισλανδίας,<sup>21</sup> με τη σύμπραξη του κράτους και ιδιωτικής εταιρείας, προκαλώντας αντιδράσεις. Κατά το σύστημα της «εικαζόμενης συναίνεσης», οι δότες των δεδομένων υγείας θεωρείται ότι σιωπηρά έχουν συναινέσει για κάθε μελλοντική χρήση, διατηρώντας το δικαίωμα να αποσύρουν τα δεδομένα τους από τη συλλογή όποτε το θελήσουν (opt out). Έτσι, θεωρείται ότι διευκολύνεται η διασύνδεση αρχείων και η επεξεργασία των μεγάλων δεδομένων, χωρίς να θίγεται η αυτονομία του δότη και ο ατομικός

έλεγχος των ευαίσθητων πληροφοριών του. Το αντεπιχείρημα στην εικαζόμενη συναίνεση προβάλλει την αξία της προηγούμενης έκφρασης της βούλησης του υποκειμένου ύστερα από κατάλληλη πληροφόρηση, για τη χρήση με συγκεκριμένο σκοπό των δεδομένων του, καθώς μόνον έτσι διασφαλίζονται όροι ελεύθερης απόφασης και πραγματικού ελέγχου. Ο λόγος είναι, ότι κάθε εκ των υστέρων «απόσυρση» δεν διασφαλίζει τον έλεγχο της χρήσης στο ενδιάμεσο χρονικό διάστημα, ενώ εξαρτάται επί πλέον από την πρωτοβουλία του ίδιου του δότη να πληροφορηθεί για πιθανές χρήσεις που ήδη έχουν δρομολογηθεί.

Η δεύτερη εναλλακτική επιλογή είναι η εγκατάλειψη της απαίτησης για «ειδική» προηγούμενη συναίνεση του δότη, και η αποδοχή μιας συναίνεσης ευρύτερης μορφής (που θα καλύπτει π.χ. περισσότερες πιθανές χρήσεις για το μέλλον). Στην εκδοχή αυτή, το πλεονέκτημα είναι η διατήρηση του όρου της προηγούμενης συναίνεσης ύστερα από πληροφόρηση –άρα του πραγματικού ελέγχου του δότη. Από την άλλη πλευρά, μειονέκτημα αποτελεί τόσο η ασάφεια του εύρους της συναίνεσης, όσο και η πιθανή εξαίρεση μελλοντικών χρήσεων που δεν μπορούν να προβλεφθούν τη στιγμή της συναίνεσης.

Ζήτημα ως προς τη συναίνεση δημιουργείται και από το ενδεχόμενο «τυχαίων ευρημάτων», κατά τη δευτερογενή επεξεργασία δεδομένων για ερευνητικούς σκοπούς. Ερωτάται, συγκεκριμένα, αν οι ερευνητές πρέπει να ενημερώνουν τους δότες των δεδομένων για ευρήματα που προκύπτουν στην πορεία της έρευνας και ενδέχεται να είναι σημαντικά για την προστασία της υγείας των τελευταίων, ή αν χρειάζεται να αναζητούν ειδική συναίνεση προς τούτο, παίρνοντας υπ' όψη το «δικαίωμα της άγνοιας».

Το πρόβλημα είναι γνωστό (και έχει αντιμετωπισθεί ειδικά στη σχετική **Γνώμη** της ΕΕΒ), ωστόσο στην περίπτωση των μεγάλων συλλογών δεδομένων, η πιθανότητα ανακάλυψης τυχαίων ευρημάτων αυξάνει σημαντικά και μάλιστα μπορεί να επεκτείνεται και σε ανώνυμα αρχικά δεδομένα, τα οποία όμως δεν αποκλείεται να ταυτοποιηθούν (βλ. αμέσως παρακάτω).

### **3.2. Η διασφάλιση της εμπιστευτικότητας των δεδομένων**

Ο μεγάλος όγκος δεδομένων που προκύπτει από τη διασύνδεση περισσότερων αρχείων θέτει ζήτημα ως προς τη διατήρηση της εμπιστευτικής χρήσης τους. Υπάρχουν εδώ δύο συγκεκριμένα προβλήματα.

Το πρώτο πρόβλημα σχετίζεται με την αναλογική αύξηση της πιθανότητας διαρροών όσο αυξάνει το μέγεθος της συλλογής. Η διαχείριση της τελευταίας απαιτεί συχνά την εμπλοκή περισσότερων υπευθύνων (τεχνικών και διοικητικών) και προϋποθέτει μεγάλη πολυπλοκότητα στα συστήματα διασφάλισης της εμπιστευτικότητας. Το στοιχείο αυτό, από μόνο του, αυξάνει τους κινδύνους παραβίασης της τελευταίας.

Το δεύτερο πρόβλημα σχετίζεται με τη σχετικοποίηση της ελεύθερης χρήσης ανώνυμων δεδομένων, στην περίπτωση που έχουμε μεγάλες συλλογές. Τα ανώνυμα δεδομένα -δηλαδή αυτά για τα οποία δεν είναι γνωστή στον επεξεργαστή η ταυτότητα του δότη, ούτε ο κώδικας της ανωνυμοποίησής τους- θεωρούνται σήμερα απόλυτα ασφαλή από άποψη εμπιστευτικότητας, καθώς δεν είναι τεχνικά δυνατή η συσχέτιση με άλλα προσωπικά δεδομένα. Το στοιχείο αυτό φαίνεται να ανατρέπεται σε συλλογές μεγάλων δεδομένων, καθώς οι σύγχρονες τεχνολογίες «βαθιάς εξόρυξης» (deep mining) οδηγούν, υπό προϋποθέσεις, ακόμη και στην ταυτοποίηση ανώνυμων δεδομένων, καθώς ο όγκος της συλλογής παρέχει πολλές παραμέτρους που διευκολύνουν συσχετίσεις για τον συγκεκριμένο σκοπό. Σε τελική ανάλυση, τα μεγάλα δεδομένα δείχνουν να ακυρώνουν την κατηγορία των ανώνυμων δεδομένων, θέτοντας το πρόβλημα της επέκτασης των συστημάτων διασφάλισης της εμπιστευτικότητας και σε αυτά τα τελευταία.

### **4. Η σχετική νομοθεσία**

Ειδική διεθνή, ενωσιακή ή εθνική νομοθεσία για τη ρύθμιση της συγκρότησης και λειτουργίας βάσεων μεγάλων δεδομένων δεν έχουμε, προς το παρόν. Ούτε η ισχύουσα νομοθεσία για την προστασία των προσωπικών δεδομένων

(ν. 2472/1997), ούτε όμως και ο νέος Κανονισμός της ΕΕ 2016/679 (που πρόκειται σύντομα να τεθεί σε εφαρμογή)<sup>22</sup> περιλαμβάνουν ιδιαίτερες προβλέψεις.

Κατ' αρχήν, επομένως, σε ότι αφορά τη συγκρότηση τέτοιων βάσεων με περιεχόμενο δεδομένα υγείας, εφαρμόζονται οι διατάξεις για την προστασία των ευαίσθητων προσωπικών δεδομένων. Σύμφωνα με αυτές (άρθ. 7 του ν. 2472/1997), η διασύνδεση αρχείων δεδομένων υγείας:

- Προϋποθέτει άδεια της Αρχής Προστασίας Δεδομένων.
- Δεν προϋποθέτει οπωσδήποτε συναίνεση των υποκειμένων.
- Αναγκαίος όρος, αν ο σκοπός της διασύνδεσης αφορά θέματα υγείας είναι η επεξεργασία να εκτελείται από πρόσωπο που υπόκειται σε καθήκον εχεμύθειας ή σε συναφή κώδικα δεοντολογίας και,
- αναγκαίος όρος, αν ο σκοπός της διασύνδεσης αφορά θέματα έρευνας είναι η ανωνυμία των δεδομένων.

Πρέπει να σημειωθεί, ότι η επέκταση του ιατρικού απορρήτου σε τρίτα πρόσωπα που υποστηρίζουν τον ιατρό και τελούν υπό την εποπτεία του (κατά το άρθ. 13 παρ. 2 α του ν. 3418/2015) δεν μπορεί να βρει εφαρμογή στην περίπτωση των βάσεων μεγάλων δεδομένων. Ο λόγος είναι ότι η συλλογή και επεξεργασία των δεδομένων, κατ' αρχήν, διενεργείται από επιστημονικό, τεχνικό και διοικητικό προσωπικό που δεν έχει την ιδιότητα του θεράποντα ως προς συγκεκριμένο υποκείμενο δεδομένων υγείας. Αυτό σημαίνει ότι η συναίνεση των υποκειμένων των δεδομένων για τη διασύνδεση των αρχείων που τα περιέχουν, παραμένει αναγκαίος όρος που συνοδεύει αυτόν της προηγούμενης άδειας της Αρχής.

Ο νέος Κανονισμός της ΕΕ δημιουργεί, ωστόσο, μια νέα εικόνα όσον αφορά τη χρήση δεδομένων στην έρευνα, που επηρεάζει καθοριστικά τη συγκρότηση βάσεων μεγάλων δεδομένων για ερευνητικούς σκοπούς στον τομέα της βιοϊατρικής. Πιο συγκεκριμένα, ο Κανονισμός δεν θεωρεί αναγκαία προϋπόθεση την ειδική προηγούμενη συναίνεση του υποκειμένου των δεδομένων, αλλά ούτε και την ανωνυμοποιημένη επεξεργασία τους, σε αντίθεση με ό,τι προβλέπει σήμερα η ισχύουσα Οδηγία 95/44. Η συναφής διάταξη του άρθρου 9 του Κανονισμού δεν

αφήνει αμφιβολίες για την αλλαγή αυτή, αν και στο προοίμιο του κειμένου του υπάρχει προβληματισμός για το ζήτημα της συναίνεσης και φαίνεται να προκρίνεται μια γενική μορφή (προηγούμενης) συναίνεσης αντί της ειδικής που γνωρίζαμε έως τώρα, ώστε σε εξαιρετικές περιπτώσεις να μη δυσχεραίνεται η έρευνα.

Η νέα αυτή ρύθμιση αντισταθμίζεται στον Κανονισμό, αφ' ενός από το δικαίωμα του προσώπου να αποσύρει τα δεδομένα του (opt-out), βάσει του γενικού «δικαιώματος λήθης» και, αφ' ετέρου, από τη μετάθεση της ευθύνης για την ασφαλή επεξεργασία των δεδομένων υπό συνθήκες εμπιστευτικότητας στον υπεύθυνο επεξεργασίας και τον νέο επιβλέποντα υπεύθυνο για την προστασία δεδομένων (Data Protection Officer). Επί πλέον, ο Κανονισμός μνημονεύει ρητά τη δυνατότητα των κρατών μελών να προβλέπουν αυστηρότερες ρυθμίσεις υπέρ της προστασίας των δεδομένων.

Πρακτικά, η ρύθμιση αυτή διευκολύνει ασφαλώς τη χρήση δεδομένων στην έρευνα -ιδίως τη «δευτερογενή» χρήση τους σε προγράμματα άλλα από εκείνα για τα οποία είχαν αρχικά συλλεγεί-, καθώς και τη διαχείριση τυχαίων ευρημάτων, δηλαδή τη δυνατότητα ενημέρωσης του υποκειμένου των δεδομένων οποτεδήποτε, καθώς η ανωνυμοποίησή τους δεν είναι πλέον αναγκαίος όρος. Από την άλλη πλευρά, ωστόσο, ο αυτοπρόσωπος έλεγχος των δεδομένων από τον φορέα τους περιορίζεται ουσιαστικά στη δυνατότητα απόσυρσής τους, κάτι που έχει ελάχιστη πρακτική σημασία, αφού είναι δύσκολο να πληροφορείται κανείς τη διαδοχική χρήση σε ποικίλα ερευνητικά προγράμματα, αν προηγουμένως δεν του έχει ζητηθεί σχετική συναίνεση.

Η πλήρης αποξένωση του υποκειμένου από τον αυτοπρόσωπο προηγούμενο έλεγχο της χρήσης ευαίσθητων δεδομένων υγείας του δημιουργεί ασφαλώς βιοηθικό πρόβλημα, στο μέτρο που θίγει τον πυρήνα της προσωπικότητας, αφού το υποκείμενο δεν έχει ουσιαστικά την ευκαιρία να αντιταχθεί στη χρήση δεδομένων του για έρευνα με τον σκοπό της οποίας ενδέχεται να διαφωνεί (π.χ. για πολεμικούς σκοπούς κ.λπ.). Μόνη λύση, ως προς τούτο, είναι οι επί μέρους εθνικοί νομοθέτες

να επιμείνουν σε πρόσθετες εγγυήσεις προστασίας, όπως η προηγούμενη συναίνεση ή η ανωνυμοποίηση των ευαίσθητων δεδομένων.

Το πρόβλημα περιπλέκεται, αν οι εθνικές νομοθεσίες ακολουθήσουν αποκλίνουσες πολιτικές εν προκειμένω, καθώς οι αυστηρότερες εθνικές ρυθμίσεις ενδέχεται να αποθαρρύνουν τη συμμετοχή σε διεθνή ερευνητικά προγράμματα, εφόσον η διακίνηση ευαίσθητων δεδομένων αντικειμενικά θα δυσκολεύεται. Επομένως, ο εθνικός νομοθέτης πρέπει να κινηθεί, εδώ, αναζητώντας μια ισόρροπη προστασία της προσωπικότητας με την ελευθερία της έρευνας.

## **5. Προτάσεις**

Η τεχνολογία των μεγάλων δεδομένων, είτε πρόκειται για δεδομένα υγείας είτε για κοινωνικά ή οικονομικά δεδομένα, αφορά τον πυρήνα της προσωπικότητας. Για τον λόγο αυτό οι φορείς ή τα μεμονωμένα άτομα (π.χ. ερευνητές) που εμπλέκονται στη διαδικασία έχουν ηθική υποχρέωση να αναγνωρίσουν του πιθανούς κινδύνους και να λάβουν τα κατάλληλα μέτρα προκειμένου να ελαχιστοποιηθεί η πιθανή βλάβη στα πρόσωπα των οποίων τα δεδομένα αναλύονται. Για την υπεύθυνη χρήση της τεχνολογίας των μεγάλων δεδομένων, στο σύνολό τους οι εμπλεκόμενοι θα πρέπει αρχικά να συνειδητοποιήσουν ότι τα δεδομένα που χειρίζονται είναι δυνατόν να έχουν αντίκτυπο στους ανθρώπους στους οποίους αναφέρονται. Στη συνέχεια, θα πρέπει να ληφθούν ειδικά μέτρα διαφύλαξης του απορρήτου των πληροφοριών.

### **5.1. Πολιτική των φορέων που κάνουν χρήση της τεχνολογίας μεγάλων δεδομένων**

Απαραίτητο βήμα πριν τη συλλογή και επεξεργασία μεγάλων δεδομένων υγείας από οργανισμούς, φορείς, εκπαιδευτικά ή ερευνητικά ιδρύματα, είναι η δημιουργία Επιτροπής Δεοντολογίας και Έρευνας με αρμοδιότητα να ελέγχει τη δεοντολογία και να εξετάζει τις αιτήσεις εγκρίνοντας ή όχι την πρόσβαση σε μεγάλα σύνολα δεδομένων και την ανάλυσή τους.<sup>23</sup> Απαραίτητη κρίνεται, επίσης, η

δημιουργία ενός Κώδικα Δεοντολογίας για τη συλλογή, πρόσβαση, διαχείριση και επεξεργασία των μεγάλων δεδομένων. Μεταξύ άλλων, ο Κώδικας Δεοντολογίας θα πρέπει να προβλέπει τον ορισμό ενός υπεύθυνου προστασίας των δεδομένων (Data Protection Officer), με αρμοδιότητα να διασφαλίζει ότι ο φορέας συμμορφώνεται με τη νομοθεσία περί προστασίας των δεδομένων.

## **5.2. Μέτρα προστασίας των δεδομένων**

Η χρήση τεχνικών μεθόδων για την προστασία των δεδομένων κρίνεται απολύτως απαραίτητη. Τα ελάχιστα μέτρα για την προστασία των δεδομένων πριν την ανάλυσή τους είναι η ανωνυμοποίησή τους με την αφαίρεση στοιχείων από εγγραφές και μητρώα, όπως ονοματεπώνυμο, διευθύνσεις κατοικίας, ημερομηνία γέννησης, αριθμός τηλεφώνου, e-mail, Αριθμός Μητρώου Κοινωνικής Ασφάλισης (ΑΜΚΑ) κτλ. Συχνά, για την αποφυγή από-ταυτοποίησης των προσώπων χρησιμοποιείται η λεγόμενη «κ ανωνυμία»,<sup>24</sup> που προσδίδει την ιδιότητα σε κάθε αναζήτηση των δεδομένων οι απαντήσεις να περιορίζονται σε κ ανθρώπους, με τη βοήθεια αλγορίθμων. Παράλληλα, εισάγεται θόρυβος στα δεδομένα και κρυπτογράφηση με αλγορίθμους, τεχνικές που, συνδυαστικά, εξασφαλίζουν ένα σχετικά καλό επίπεδο ασφάλειας.

## **5.3. Συγκατάθεση ύστερα από ενημέρωση**

Τα μη ανωνυμοποιημένα δεδομένα θα πρέπει να αναλύονται μετά από συγκατάθεση των προσώπων που αφορούν και τα έντυπα συγκατάθεσης θα πρέπει να περιλαμβάνουν πληροφορίες σχετικά με:

1. Την αποθήκευση των δεδομένων, ποιος θα έχει πρόσβαση σε αυτά και τι μέτρα θα ληφθούν για να εμποδιστεί η πρόσβαση από τρίτους.
2. Την πιθανή διακίνηση των δεδομένων μεταξύ φορέων ή ερευνητών στην ίδια ή σε άλλες χώρες, καθώς και το εάν θα ανταλλάσσονται τα δεδομένα στην αρχική τους μορφή πριν την επεξεργασία ή μόνο τα αποτελέσματα της ανάλυσής τους.



3. Την πιθανότητα παραγωγής τυχαίων ευρημάτων που αφορούν στην υγεία των προσώπων όπου αναφέρονται τα δεδομένα, καθώς και την πολιτική που θα ακολουθηθεί στις περιπτώσεις αυτές. Εάν, δηλαδή, τα τυχαία ευρήματα θα γνωστοποιηθούν στους ενδιαφερόμενους, καθώς και ποια θα είναι αυτά (π.χ. μόνο όσα έχουν συγκεκριμένη κλινική σημασία).<sup>25</sup>
4. Την πιθανότητα εμπορικής χρήσης των δεδομένων.

## ΒΙΒΛΙΟΓΡΑΦΙΑ

- <sup>1</sup> Berkman BE, Shapiro ZE, Eckstein L, Pike ER. The Ethics of Large-Scale Genomic Research. In J. Collmann and S.A. Matei (eds.), *Ethical Reasoning in Big Data, Computational Social Sciences*, Springer International Publishing Switzerland 2016, pp 53-54.
- <sup>2</sup> Βιδάλης Τ, Μολλάκη Β. Ηλεκτρονικός Φάκελος Υγείας: Ζητήματα Δεοντολογίας (2015). Έκθεση για την Εθνική Επιτροπή Βιοηθικής. <http://www.bioethics.gr/index.php/el/gnomes/910-2015-04-28-08-33-51>
- <sup>3</sup> Mittelstadt BD, Fairweather NB, Shaw M, McBride N. The ethical implications of personal health monitoring. *International Journal of Technoethics* 2014, 5(2), 37–60.
- <sup>4</sup> Boye N. Co-production of Health enabled by next generation personal health systems. *Studies in health technology and informatics* 2012, 177, 52–58.
- <sup>5</sup> Ginsberg J, Mohebbi M, Patel R, Brammer L. Detecting influenza epidemics using search engine query data. *Nature* 2009, 457, 1012–1014.
- <sup>6</sup> Ortiz J, Zhou H, Shay D, Neuzil K, Fowlkes A. Monitoring influenza activity in the United States: a comparison of traditional surveillance systems with Google Flu Trends. *PLoS One* 2011, 6, e18687.
- <sup>7</sup> Broniatowski DA, Paul MJ, Dredze M. Letter “Twitter: Big Data Opportunities,” *Science* 2014, 345, 6193: 148.
- <sup>8</sup> Kramer ADI., Guillory JE, Hancock JT. Experimental Evidence of Massive Scale Emotional Contagion through Social Networks. *Proceedings of the National Academy of Sciences U.S.A.* 2014, 111, 24 (2014): 8788-8790.
- <sup>9</sup> Halamka J. Early experiences with big data at an academic medical center. *Health Affairs* 2014, 33(7): 1132–1138.
- <sup>10</sup> Kohane IS. Using electronic health records to drive discovery in disease genomics. *Nat Rev Genet* 2011 Jun;12(6):417-28.
- <sup>11</sup> Xu H, Fu Z, Shah A, Chen Y, Peterson NB, Chen Q, Mani S, Levy MA, Dai Q, Denny JC. Extracting and integrating data from entire electronic health records for detecting colorectal cancer cases. *AMIA Annu Symp Proc* 2011;2011:1564-72.
- <sup>12</sup> Weiss J, Natarajan S, Paissig P, McCarthy C, Page D. Machine learning for personalized medicine: Predicting primary MI from electronic medical records. *AI Magazine* 2012, 33(4): 33-45.
- <sup>13</sup> Li L, Lu J, Xue W, Wang L, Zhai Y, Fan Z, Wu G, Fan F, Li J, Zhang C, Zhang Y, Zhao J. Target of obstructive sleep apnea syndrome merge lung cancer: based on big data platform. *Oncotarget*. 2017 Mar 28;8(13):21567-21578.
- <sup>14</sup> Zhang X, Pérez-Stable EJ, Bourne PE, Peprah E, Duru OK, Breen N, Berrigan D, Wood F, Jackson JS, Wong DWS, Denny J. Big Data Science: Opportunities and Challenges to Address Minority Health and Health Disparities in the 21st Century. *Ethn Dis* 2017 Apr 20;27(2):95-106.
- <sup>15</sup> Lipworth W, Mason PH, Kerridge I, Ioannidis JP. Ethics and Epistemology in Big Data Research. *J Bioeth Inq*. 2017 Mar 20.
- <sup>16</sup> El Emam K, Jonker E, Arbuckle L, Malin B. A systematic review of re-identification attacks on health data. *PLoS One* 2011, 6(12):e28071.

---

<sup>17</sup> Barbaro M, Zeller T, Hansell S. A face is exposed for AOL searcher no. 4417749. New York Times 2006 Aug 9;9.

<sup>18</sup> Cox J. 70,000 OkCupid Users Just Had Their Data Published. Motherboard. 12 May 2016. <http://motherboard.vice.com/read/70000-okcupid-users-just-had-their-data-published>

<sup>19</sup> Sweeney L. k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. 2002, 10(05):557–70.

<sup>20</sup> Kloumann IM, Kleinberg JM. Community membership identification from small seed sets. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining 2014 Aug 24 (pp. 1366–1375). ACM.

<sup>21</sup> Gottweis H, Zatloukal K. Biobank Governance: Trends and Perspectives, Pathobiology 2007,74:206–211, <https://www.karger.com/Article/Pdf/104446>

<sup>22</sup> Μήτρου Λ. Ο Γενικός Κανονισμός Προστασίας Προσωπικών Δεδομένων, εκδ. Σάκκουλα 2017, Αθήνα - Θεσσαλονίκη.

<sup>23</sup> Μολλάκη Β., Βιδάλης Τ. Καταγραφή του υφιστάμενου καθεστώτος αναφορικά με την ίδρυση και λειτουργία των Επιτροπών Δεοντολογίας της Έρευνας: Συνάντηση με εκπροσώπους των Α.Ε.Ι. και των Ερευνητικών Ιδρυμάτων (2015). Έκθεση για την Εθνική Επιτροπή Βιοηθικής. [http://www.bioethics.gr/images/pdf/GNOMES/REPORT\\_ETHICS\\_COMMITTEES\\_FINAL\\_GR.pdf](http://www.bioethics.gr/images/pdf/GNOMES/REPORT_ETHICS_COMMITTEES_FINAL_GR.pdf)

<sup>24</sup> Latanya S., Pierangela S. "Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression". Harvard Data Privacy Lab 1998.

<sup>25</sup> Μολλάκη Β, Βιδάλης Τ. Τυχαία Ευρήματα στην Έρευνα και την Κλινική Πράξη (2015). Έκθεση για την Εθνική Επιτροπή Βιοηθικής. [http://www.bioethics.gr/images/pdf/GNOMES/REPORT\\_INCIDENTAL\\_FINDINGS\\_FINAL\\_GR.pdf](http://www.bioethics.gr/images/pdf/GNOMES/REPORT_INCIDENTAL_FINDINGS_FINAL_GR.pdf)