



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Η χρήση των Big Data Analytics για τη βελτίωση των ψηφιακών υπηρεσιών υγείας

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Τσώλας Λεωνίδας

Επιβλέπων : Διονύσιος-Δημήτριος Κουτσούρης

Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούνιος 2017



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Η χρήση των **Big Data Analytics** για τη βελτίωση των ψηφιακών υπηρεσιών υγείας

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Τσώλας Λεωνίδας

Επιβλέπων : Διονύσιος-Δημήτριος Κουτσούρης

Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 28^η Ιουνίου 2017.

.....
Διονύσιος-Δημήτριος
Κουτσούρης

Καθηγητής Ε.Μ.Π.

.....
Παναγιώτης Τσανάκας

Καθηγητής Ε.Μ.Π.

.....
Γεώργιος Ματσόπουλος

Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούνιος 2017

.....

Τσώλας Λεωνίδα

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Τσώλας Λεωνίδα, 2017

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ	9
ABSTRACT	10
ΚΕΦΑΛΑΙΟ 1 - ΕΙΣΑΓΩΓΗ	13
1.1 Πρόλογος.....	14
1.2 Αντικείμενο της διπλωματικής.....	15
ΚΕΦΑΛΑΙΟ 2 - Η ΕΝΝΟΙΑ ΤΩΝ BIG DATA ΣΤΟΝ ΥΓΕΙΟΝΟΜΙΚΟ ΤΟΜΕΑ	17
2.1 Ορισμός και Χαρακτηριστικά Big Data.....	19
2.2 Πηγές δεδομένων στην υγεία και μεθοδολογία.....	31
2.3 Πεδία Εφαρμογής.....	34
ΚΕΦΑΛΑΙΟ 3 - ΕΡΓΑΛΕΙΑ ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΓΙΑ ΤΗΝ ΑΞΙΟΠΟΙΗΣΗ ΤΩΝ BIG DATA	39
3.1 Βασικά εργαλεία.....	41
3.2 Επιπρόσθετα εργαλεία.....	46
3.3 Εμπορικές εφαρμογές.....	58
ΚΕΦΑΛΑΙΟ 4 - ΣΥΜΒΟΛΗ ΣΤΗΝ ΥΓΕΙΑ	61
4.1 Πεδία δράσης και ευκαιρίες.....	63
4.3 Ηλεκτρονικός Φάκελος Υγείας (EHR).....	70
4.4 Οι ευκαιρίες με το μεγαλύτερο αντίκτυπο.....	74
4.5 Ερευνητικό έργο και δράσεις.....	80
4.6 Ιστορίες επιτυχίας στη φροντίδα ασθενών.....	100
ΚΕΦΑΛΑΙΟ 5 - ΤΕΧΝΙΚΕΣ ΚΑΙ ΟΡΓΑΝΩΤΙΚΕΣ ΠΡΟΚΛΗΣΕΙΣ	112
5.1 Αποθετήρια Δεδομένων (Data Silos).....	114
5.2 Πολιτικές, κοινωνικές και βιοηθικές επιπτώσεις.....	116
5.3 Το ευρύτερο πεδίο των προκλήσεων.....	120
5.4 Οι νέες τάσεις των Big Data στην υγεία.....	124
ΚΕΦΑΛΑΙΟ 6 - ΕΠΙΛΟΓΟΣ	128
6.1 Συμπεράσματα.....	130
6.2 Επίλογος.....	132
ΒΙΒΛΙΟΓΡΑΦΙΑ	135

ΠΕΡΙΛΗΨΗ

Η βασική επιδίωξη των συστημάτων υγείας είναι η παροχή υπηρεσιών υγειονομικής περίθαλψης σε υψηλή ποιότητα, με χαμηλό κόστος και καθολική προσβασιμότητα. Ο ανταγωνισμός μεταξύ των χαρακτηριστικών αυτών, δημιουργεί την ανάγκη ριζικών αλλαγών ικανών να ανατρέψουν τις ισορροπίες. Παρόλο που υπάρχει διαθέσιμη πληροφορία και γνώση σε βαθμό που να επιτρέπει βελτιώσεις σε πολλούς τομείς των υπηρεσιών υγείας, παρατηρείται εγγενής αδυναμία εξόρυξης συμπερασμάτων σε μεγάλη κλίμακα, προσαρμοσμένων στις επιμέρους απαιτήσεις. Το κενό αυτό καλύπτεται αποτελεσματικά μέσω της ανάλυσης μεγάλων συνόλων δεδομένων ετερογενών πηγών. Στην παρούσα διπλωματική εργασία περιγράφονται τα χαρακτηριστικά των δεδομένων του τομέα της υγειονομικής περίθαλψης, οι ιδιαιτερότητές τους και τα διαθέσιμα εργαλεία για την αξιοποίησή τους. Επιπλέον, δίδεται ιδιαίτερη έμφαση στο υπό εξέλιξη ερευνητικό έργο και τα αποτελέσματα που έχουν επιτευχθεί. Μελετώνται οι τεχνικές και οργανωτικές προκλήσεις και οι μελλοντικές προοπτικές. Η εργασία εκπονήθηκε με στόχο την πλήρη παρουσίαση όλων των χαρακτηριστικών που οφείλουν να ληφθούν υπόψη από οργανισμούς και φορείς υγείας που αποσκοπούν στην αξιοποίηση των «μεγάλων δεδομένων». Στο πλαίσιο αυτό, παρουσιάζεται η προτεινόμενη μεθοδολογική προσέγγιση και καθίσταται εφικτή η αποτελεσματική στοχοθεσία, η πρόβλεψη των ζητημάτων ενδέχεται να προκύψουν, καθώς και ο αναμενόμενος αντίκτυπος.

Λέξεις-κλειδιά: Big Data, υγειονομική περίθαλψη, Analytics, Hadoop, Spark, IBM Watson, Ηλεκτρονικός Φάκελος Υγείας, Δεδομένα, Βιοϊατρική, Τεχνολογία, Υγεία, EMIF, OpenPHACTS, Silos.

ABSTRACT

The basic aim of healthcare systems is to provide high quality, low-cost and universal access healthcare services. The competition between these features creates the need for radical changes capable of changing the balance. Although there is enough information and knowledge to allow improvements in many areas of healthcare services, there is inherent inability to draw large-scale conclusions tailored to individual requirements. This gap is effectively covered by the analysis of large heterogenous data sets from different sources. This diploma thesis describes the characteristics of healthcare data, their specificities and the available tools for their exploitation. Moreover, special emphasis is given to the ongoing research work and the results achieved. This study includes the technical and organizational challenges and the future prospects. It is designed to fully present all the features that should be taken into account by healthcare organizations aiming at exploiting Big Data. Therefore, it presents the proposed methodological approach that enables effective targeting, prediction of the issues that may arise, and the expected impact.

Key-words: Big Data, Healthcare, Analytics, Hadoop, Spark, IBM Watson, Electronic Health Records, Data, Biomedicine, Technology, Health, EMIF, OpenPHACTS, Silos.

Κατάλογος ακρωνυμίων

API	Application Programming Interface
AWS	Amazon Web Services
CDC	Center for Disease Control
EBI	European Bioinformatics Institute
EHR	Electrical Health Record
EMIF	European Medical Information Framework
EMIF-AD	EMIF-Alzheimer Disease
EMR	Electrical Medical Record
ESRI	Environmental Systems Research Institute
FDA	Food and Drug Administration
HDFS	Hadoop Distributed File System
HIMSS	Healthcare Information and Management Systems Society
HIPAA	Health Insurance Portability and Accountability Act
ICU	Intensive Care Units
IMI	Innovative Medicines Initiative
IOT	Internet of Things
JDBC	Java Database Connectivity
JSON	Java Script Object Notation
MDM	Master Data Management
MLlib	Machine Learning Library
NIH	National Institute of Health
NLP	Natural Language Processing
ODBC	Open Database Connectivity
OMOP	Observational Medical Outcomes Partnership
OpenPHACTS	Open PHArmacological Concept Triple Store
RDD	Resilient Distributed Dataset
SQL	Structured Query Language
WHO	World Health Organization

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

1.1 Πρόλογος

Στα πλαίσια της Ιατρικής επιστήμης συνεχώς δημιουργούνται ανάγκες που σηματοδοτούν εξελίξεις στον επιστημονικό τομέα, ενώ, παράλληλα, οι καινοτομίες και οι νέες τεχνολογίες βρίσκουν ευρεία εφαρμογή στον κλάδο της υγείας. Η τεχνολογική εξέλιξη των τελευταίων χρόνων έχει προκαλέσει ραγδαία αύξηση του συνόλου των διαθέσιμων δεδομένων. Ενδεικτικά, στις Ηνωμένες Πολιτείες το 2011, το σύστημα υγείας είχε «πληροφοριακό μέγεθος» 150 exabytes (10^{18} bytes).¹

Από τη συνεχή αυτή σχέση αλληλεπίδρασης Ιατρικής – Τεχνολογίας δεν θα μπορούσε να εξαιρεθεί το φαινόμενο της χρήσης και αξιοποίησης των μεγάλων αυτών όγκων δεδομένων για τα οποία έχει επικρατήσει ο όρος Big Data. Η αέναη ανάγκη τελειοποίησης των ιατρικών μεθόδων για έγκαιρη πρόγνωση και διάγνωση ασθενειών, καθώς και η εφαρμογή αποτελεσματικών μεθόδων θεραπείας και αποκατάστασης προϋποθέτει την ενδελεχή εξέταση σχέσεων αιτίου-αποτελέσματος και την παραγωγή κατάλληλων μοντέλων. Η Ιατρική συνεχώς δημιουργεί νέα απεικονιστικά δεδομένα, δεδομένα από βασική και κλινική έρευνα και επιδημιολογία, από διοικητικές υπηρεσίες υγείας και ασφαλιστικούς οργανισμούς, από υπηρεσίες δημόσιας υγείας αλλά και από μη συμβατικές πηγές δεδομένων όπως κοινωνικά δίκτυα, εφαρμογές διαδικτύου κ.ό.κ. Ο κλινικός ιατρός αποκτά πολλαπλά οφέλη από την αξιοποίηση των Big Data, καθώς χρησιμεύουν ως νέα εργαλεία υποβοήθησης στη λήψη αποφάσεων, στη βελτίωση μεθοδολογιών κλινικής έρευνας και της αποτελεσματικότητας των θεραπειών, αλλά και στην εξατομικευμένη Ιατρική. Τέλος, υπάρχει μεγάλο όφελος στην εξοικονόμηση πόρων και ανθρώπινου δυναμικού και στην ανακατανομή τους για την αύξηση της παραγωγικότητας και τον εξορθολογισμό των δαπανών.

Οι επαγγελματίες στο χώρο της υγείας αντιλαμβάνονται τις ριζικές αλλαγές που μπορεί να επιφέρει η χρήση των Big Data στον τομέα της υγειονομικής περίθαλψης και για πρώτη φορά σε τέτοια κλίμακα η δυνατότητα εξαγωγής συμπερασμάτων από τεράστιο όγκο σύνθετων και ετερογενών δεδομένων σε μικρό χρονικό διάστημα, δίχως την ανάγκη δαπάνης υπέρογκων χρηματικών ποσών. Συνεπώς η χρήση των Big Data, με όλα τα πλεονεκτήματα και τις νέες δυνατότητες που παρέχει, την ευρεία χρήση που ήδη γνωρίζει, επιβεβαιώνεται πως δεν αποτελεί μία ακόμα φέρελπη νέα τεχνολογία που προορίζεται για το μέλλον. Έχει ήδη δώσει τα απαραίτητα εχέγγυα πως αποτελεί το παρόν της Ιατρικής επιστήμης.

1.2 Αντικείμενο της διπλωματικής

Η παρούσα διπλωματική εργασία στοχεύει στην κατανόηση του τρόπου χρήσης των Big Data στον τομέα των ψηφιακών υπηρεσιών υγείας, στην παρουσίαση των διαθέσιμων εργαλείων και εφαρμογών, στην περιγραφή των σημαντικότερων δράσεων και ερευνητικού έργου, καθώς και των τεχνικών και οργανωτικών προκλήσεων που εγείρονται. Οι οργανισμοί και φορείς υγείας που εξετάζουν την αξιοποίηση της τεχνολογίας των Big Data Analytics, καθώς και όσοι ήδη κάνουν χρήση της, έχουν τη δυνατότητα, μέσω της παρούσας να μελέτης, να αποκτήσουν εμπειριστατωμένη και ολοκληρωμένη εικόνα για τις δυνατότητες αξιοποίησης, την αποτελεσματική στόχευση, την προτεινόμενη μεθοδολογική προσέγγιση, τον αναμενόμενο αντίκτυπο και τα ζητήματα που θα αντιμετωπίσουν στα διάφορα στάδια του εγχειρήματος.

ΚΕΦΑΛΑΙΟ 2

Η ΕΝΝΟΙΑ ΤΩΝ BIG DATA ΣΤΟΝ ΥΓΕΙΟΝΟΜΙΚΟ ΤΟΜΕΑ

2.1 Ορισμός και Χαρακτηριστικά Big Data

2.1.1 Ορισμός

Δίχως να είναι γνωστό ποιος χρησιμοποίησε για πρώτη φορά τον όρο Big Data για να περιγράψει σύνολα δεδομένων, αυτός που τον καθιέρωσε ήταν ο John Mashey.^{2 3} Μάλιστα, ο όρος φαίνεται να χρησιμοποιείται ευρέως ήδη από τη δεκαετία του 1990. Ποιος είναι όμως ο τυπικός ορισμός της έννοιας;

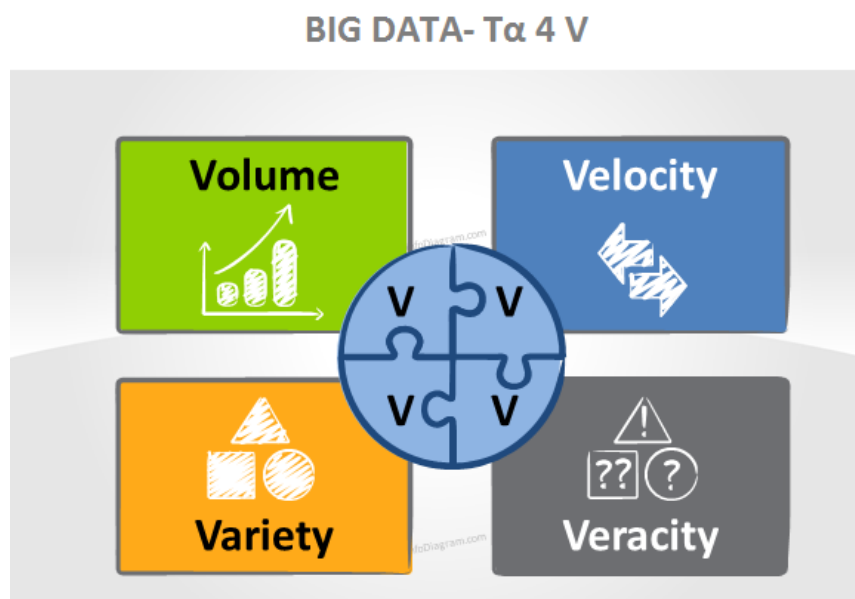
Ορισμένες πηγές του Διαδικτύου αναφέρουν ενδεικτικά ότι μία συλλογή δεδομένων χαρακτηρίζεται ως Big Data όταν είναι της τάξεως των petabytes (10^{15} bytes) ή exabytes (10^{18} bytes). Ωστόσο, οι περισσότεροι επιστήμονες συμφωνούν ότι δεν είναι απαραίτητο να ποσοτικοποιηθεί επακριβώς ο όρος Big Data, αφού μία τέτοια προσέγγιση δεν θα είχε κάποιο νόημα. Όπως μάλιστα επισημαίνει η Kate Crawford, διακεκριμένη επιστήμονας δεδομένων (data scientist), «δεν τίθεται αμφιβολία ότι το μέγεθος των διαθέσιμων δεδομένων είναι πράγματι τεράστιο, αλλά αυτό δεν είναι το σημαντικότερο χαρακτηριστικό των Big Data».⁴

Για το λόγο αυτό είθισται να θεωρείται Big Data μία συλλογή δεδομένων τόσο μεγάλη ή τόσο ετερογενής και σύνθετη ως προς τη δομή της, ώστε τα παραδοσιακά λογισμικά επεξεργασίας δεδομένων είναι αδύνατον να τη διαχειριστούν. Οι προκλήσεις που εγείρουν τα Big Data περιλαμβάνουν τη συγκέντρωση, αποθήκευση, ανάλυση, μεταφορά, διαμοιρασμό καθώς και οπτική απεικόνιση της πληροφορίας που περιέχουν. Επιστήμονες, επιχειρηματίες και επαγγελματίες στον τομέα της υγείας συχνά καλούνται να χρησιμοποιήσουν δεδομένα τα οποία δεν προέρχονται από μία μοναδική πηγή είτε από πηγές ίδιας δομής, αλλά μπορεί να προέρχονται από διαφορετικές πηγές, όπως η διεθνής βιβλιογραφία, το διαδίκτυο, ιατρικά αρχεία και μητρώα ασθενών, ακόμα και δεδομένα «έξυπνων» συσκευών. «Τα κλασικά σχεσιακά συστήματα διαχείρισης βάσεων δεδομένων αδυνατούν να επεξεργαστούν τα Big Data, καθώς απαιτείται η μεγάλης κλίμακας παράλληλη εκτέλεση λογισμικού σε δεκάδες, εκατοντάδες, ακόμα και χιλιάδες εξυπηρετητές (servers)».⁵ Αρχίζει λοιπόν να γίνεται κατανοητό πως ο όρος Big Data εξαρτάται από το υλικό (hardware) και τα εργαλεία λογισμικού (software tools) που διαθέτει ο χρήστης και δεν μπορεί να ορισθεί με απόλυτο τρόπο.

Αξίζει να αναφερθεί ότι συχνά ο όρος Big Data χρησιμοποιείται, καταχρηστικά, ταυτόσημα με τον όρο Big Data Analytics, δηλαδή με τη διαδικασία ανάλυσης των δεδομένων και την μετέπειτα εξαγωγή χρήσιμων συμπερασμάτων προσαρμοσμένων στο εκάστοτε πρόβλημα.

2.1.2 Χαρακτηριστικά Big Data

Τα χαρακτηριστικά των Big Data τα οποία εξετάζονται και χρησιμοποιούνται ως τρόποι διάκρισης και κατηγοριοποίησης τους είναι γνωστά ως “4+1 V’s” και είναι τα εξής:⁶



Εικόνα 2.1: Τα χαρακτηριστικά των 4 V

Όγκος (Volume)

Ο όγκος των παγκόσμιων δεδομένων αυξάνεται με εκθετικούς ρυθμούς. Από 130 exabytes (10^{18} bytes) δεδομένων το 2005, αυξήθηκε σε 7,910 exabytes το 2015.⁷ Μέχρι το 2020 θα υπάρχουν συνολικά 35 zettabytes (10^{21} bytes) ψηφιακών δεδομένων. Ενδεικτικά, αν αυτή ποσότητα αποθηκευόταν εξ ολοκλήρου σε δίσκους DVD, θα δημιουργούνταν μία στοίβα από DVD, η οποία θα εκτεινόταν από τη Γη ως τον Άρη!⁸ Παρ' όλα αυτά, μόνο το 20% των συνολικών δεδομένων είναι δομημένα, δηλαδή άμεσα επεξεργάσιμα από υπολογιστικά συστήματα και το υπόλοιπο 80% είναι μη δομημένα δεδομένα (χειρόγραφες σημειώσεις, κείμενα χωρίς ετικέτες, αρχείο ήχου ή βίντεο κ.ό.κ.)⁹ Ήδη, το 2016, υπήρχαν στην παγκόσμια αγορά περισσότερα από 1 δισεκατομμύριο «έξυπνα» κινητά τηλέφωνα (smartphones), 400 εκατομμύρια ψηφιακές ταμπλέτες (tablets) και 1 δισεκατομμύριο προσωπικοί ηλεκτρονικοί υπολογιστές με δυνατότητα σύνδεσης στο διαδίκτυο (Internet) και κατ' επέκταση, κοινή χρήση δεδομένων.⁷

Ποικιλία (Variety)

Πέραν του τεράστιου όγκου των διαθέσιμων δεδομένων, η ποικιλία τους, εάν δηλαδή πρόκειται για δεδομένα τα οποία είναι δομημένα (structured), μη δομημένα (unstructured) ή ημι-δομημένα (semi-structured), τα καθιστά μια ιδιαίτερα ενδιαφέρουσα πρόκληση. Τα δομημένα είναι δεδομένα που μπορούν εύκολα να αποθηκευθούν, να αναλυθούν και να αξιοποιηθούν από κάποιο υπολογιστικό σύστημα, ανεξάρτητα εάν είναι εύκολα αναγνώσιμα από τον άνθρωπο. Τα ημι-δομημένα αποτελούν μία μορφή δεδομένων με κάποια δομή, η οποία όμως δεν έρχεται σε συμφωνία με την επίσημη δομή των μοντέλων διαχείρισης δεδομένων που περιγράφουν οι σχεσιακές βάσεις δεδομένων, αλλά παρόλα αυτά, διαθέτουν ετικέτες (tags) ώστε να υπάρχει ένα είδος ιεραρχίας και σημασιολογικής διάκρισης των στοιχείων τους.¹⁰ Τα δεδομένα που δεν εμπίπτουν στις δύο αυτές κατηγορίες είναι τα μη-δομημένα, τα οποία είτε δεν έχουν κάποιο προκαθορισμένο περιγραφικό μοντέλο ή δεν είναι οργανωμένα βάση κάποιας συγκεκριμένης δομής.

Ταχύτητα (Velocity)

Η συνεχής ροή δεδομένων που συσσωρεύονται με πρωτοφανή ρυθμό παρουσιάζει, όπως είναι αναμενόμενο, νέες προκλήσεις. Όπως ακριβώς έχει μεταβληθεί ο όγκος και η ποικιλία των λαμβανόμενων δεδομένων, έτσι και η ταχύτητα με την οποία αυτά καταφθάνουν, αποθηκεύονται, επεξεργάζονται και τέλος αναλύονται ώστε να παρθούν αποφάσεις βάσει αυτών, έχει αυξηθεί ραγδαία. Η στροφή από τις επιταγές στις πιστωτικές κάρτες για τις οικονομικές δοσοληψίες αποτελεί αντιπροσωπευτικό παράδειγμα μετάβασης από μεθόδους χρονοβόρας επεξεργασίας σε μεθόδους ταχείας, αυτοματοποιημένης διαδικασίας. Αντίστοιχες προκλήσεις παρατηρούνται και στον τομέα της υγείας.

Εγκυρότητα (Veracity)

Τα παραδοσιακά συστήματα διαχείρισης δεδομένων αποδέχονται την εγκυρότητα αυτών, την ακρίβεια και την ορθότητά τους. Ωστόσο, οποιοσδήποτε έχει συναντήσει κάποια παραφωνία σε κάποια ηλεκτρονική υπηρεσία, αντιλαμβάνεται πως αυτή η παραδοχή δεν είναι πάντοτε ορθή.

Αξία (Value)

Πρόκειται για τον τελευταίο και επιπρόσθετο παράγοντα, ενδεικτικό για τα οφέλη που επιφέρει η χρήση των Big Data, και εΐθισται να μετράται τόσο με χρήση δεικτών, όσο και με στατιστικές αναλύσεις που ποσοτικοποιούν τα αποτελέσματα που προκύπτουν, ενώ αφορούν και τον οικονομικό τομέα, αλλά όχι αποκλειστικά.

Τέλος, αξίζει να αναφερθεί πως σε ορισμένες βιβλιογραφικές πηγές συναντάται ένας επιπλέον παράγοντας, η **Μεταβλητότητα (Variability)** των Big Data. Πρόκειται για συνεχείς μεταβολές των δεδομένων που λαμβάνονται και είναι πιθανόν να δυσχεράνουν την ανάλυση και αξιοποίησή τους.

Σε κάποιες βιβλιογραφικές πηγές συναντάται ο όρος **σθένος (Valence)** που είναι η αντίθετη έννοια της μεταβλητότητας, αλλά στην ουσία προσδιορίζει το ίδιο χαρακτηριστικό.

Εν ολίγοις, μία συλλογή δεδομένων, συγκαταλέγεται στην κατηγορία Big Data, εάν καλύπτει τις ακόλουθες προδιαγραφές:

- **Μεγάλος όγκος**
- **Ποικιλία, ανομοιογένεια**
- **Εγκυρότητα**
- **Δυνατότητα αποκόμισης οφέλους από την αξιοποίησή τους**

Ο παράγοντας της **ταχύτητας** δεν αναφέρθηκε σκόπιμα, καθώς Big Data μπορούν να θεωρηθούν συλλογές δεδομένων, ανεξάρτητα από το ρυθμό προσέλευσής. Για την επεξεργασία τους ακολουθείται βέβαια, διαφορετική διαδικασία.

2.1.3 Χαρακτηριστικά Big Data στο χώρο της Υγείας

Οι Frost και Sullivan είναι οι εισηγητές του όρου Advanced Health Analytics.¹¹ Είναι χρήσιμο στο σημείο αυτό να παρουσιαστούν τα βασικά χαρακτηριστικά των Big Data προσαρμοσμένα στον τομέα της ψηφιακής υγείας:¹²

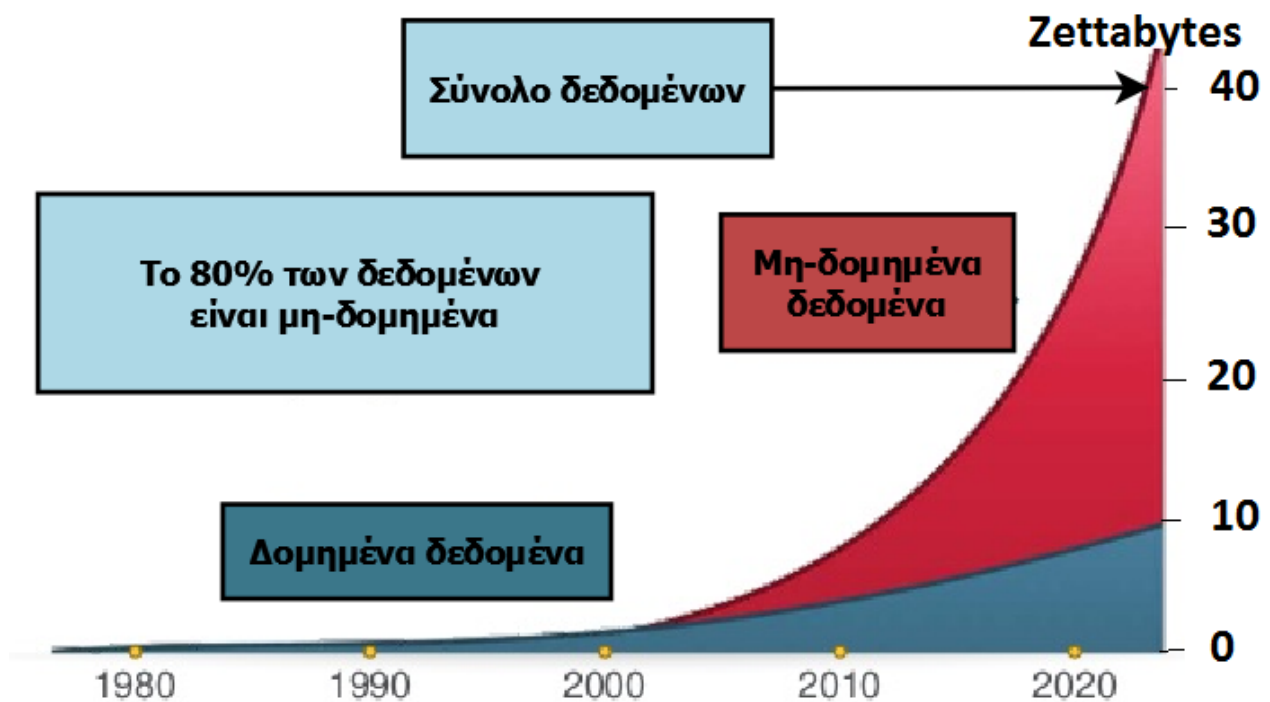
Όγκος (Volume)

Στον τομέα της ψηφιακής υγείας, η αύξηση της ποσότητας των δεδομένων είναι αποτέλεσμα τόσο ψηφιοποίησης ήδη διαθέσιμων δεδομένων, όσο και δημιουργίας νέων μορφών δεδομένων. Ο όγκος των διαθέσιμων δεδομένων αποτελείται από προσωπικά ιατρικά αρχεία, εικόνες ραδιολογίας και ακτινολογίας, κλινικές δοκιμές, έρευνες, δημογραφικά στοιχεία, ανθρώπινα γονιδιώματα, γενετικές ακολουθίες κ.ό.κ. Οι νέες μορφές Big Data, όπως εικόνες τριών διαστάσεων (3D), δεδομένα γονιδιωματικών και βιομετρικών αισθητήρων, συμβάλλουν στην εκθετική αύξηση των δεδομένων στον κλάδο της υγείας.

Ο όγκος των διαθέσιμων ιατρικών δεδομένων ήταν 500 petabytes (10^{15} bytes) το 2012 και αναμένεται να ξεπεράσει τα 25,000 petabytes έως το 2020. Οι εξελίξεις στη διαχείριση των δεδομένων, κυρίως στην οπτική απεικόνιση, δηλαδή στην αναπαράσταση των δεδομένων μέσω εικόνων, και η χρήση του υπολογιστικού νέφους (cloud computing), έχουν οδηγήσει στην διαμόρφωση εφαρμογών που στοχεύουν στην αποτελεσματικότερη λήψη, αποθήκευση και διαχείριση των Big Data. Η αποθήκευση των δεδομένων σε υπολογιστικά νέφη καθιστά εφικτή την πρόσβασή τους από διαφορετικές συσκευές άμεσα, ταχύτατα, δημιουργώντας έτσι, τελικά, ένα απέραντο σύμπαν από πληροφορίες, στο οποίο οι εμπορικές συσκευές μπορούν να έχουν πρόσβαση, χωρίς να απαιτούνται υπέρογκα χρηματικά ποσά. Αρκετές εταιρίες, κολοσσοί στην τεχνολογία, όπως η Cisco Systems Inc., η IBM, η Oracle Corporation, η Google, η Amazon και πολλές ακόμα προσπαθούν να βελτιώσουν τη διαχείριση των δεδομένων, ενώ έχουν δημιουργήσει πλατφόρμες αξιοποίησής τους, όπως θα παρουσιαστεί και στη συνέχεια.

Ποικιλία (Variety)

Παραδοσιακά, η συντριπτική πλειοψηφία των διαθέσιμων δεδομένων στον τομέα της υγείας ήταν μη δομημένα δεδομένα, όπως ιατρικές καταγραφές, χειρόγραφες σημειώσεις ιατρικού και νοσηλευτικού προσωπικού για την περιγραφή συμπτωμάτων, ενδείξεων, συμπεριφοράς, ιατρικές εικόνες κ.ό.κ. Βέβαια, τα τελευταία έτη, έχουν αυξηθεί τα δομημένα δεδομένα, όπως πληροφορίες σχετικά με ηλεκτρονική συνταγογράφηση φαρμάκων, ποσοτικά δεδομένα μετρήσεων οργάνων και εξετάσεων, και γενικότερα δεδομένα τα οποία υπάρχει προσπάθεια να καταγραφούν σε μία ενιαία δομή, ώστε να είναι αξιοποιήσιμα στη συνέχεια. Εκτός από τα δεδομένα τα οποία είναι προφανές ότι καταγράφονται, στον τομέα της υγείας είναι ιδιαίτερα χρήσιμα και δεδομένα νέων πηγών, όπως συσκευών ευεξίας που καταγράφουν τους παλμούς ή τη διάρκεια ύπνου των ασθενών, κοινωνικών δικτύων, γονιδιωματικής έρευνας κ.ό.κ. Τα δεδομένα αυτά, αν και ιδιαίτερα χρήσιμα, δεν είναι προς το παρόν αξιοποιήσιμα στο έπακρο, καθώς δεν είναι εύκολη η μετατροπή από μία μη-δομημένη σε μία δομημένη μορφή. Υπάρχουν πάντως πολλές εφαρμογές που τα αξιοποιούν και στοχεύουν στη βελτίωση της διαχείρισής τους.



Εικόνα 2.2: Ο συνολικός όγκος δεδομένων ανά έτος και η ραγδαία αύξηση των μη δομημένων δεδομένων

Τα δομημένα δεδομένα στα ηλεκτρονικά ιατρικά αρχεία (**Electrical Medical Records** ή **EMR**) και ηλεκτρονικά αρχεία υγείας (**Electrical Health Records** ή **EHR**) περιλαμβάνουν πεδία όπως το ονοματεπώνυμο του ασθενούς, η ημερομηνία γέννησης, η διεύθυνση, το όνομα του θεράποντος ιατρού, το νοσοκομείο ή η κλινική, η αντίστοιχη διεύθυνση κ.ό.κ. Η ομαδοποίηση των δεδομένων σε πεδία, παρά το αδιαμφισβήτητο όφελος και τον περιορισμό των λαθών που μπορεί να προσφέρει, συναντά τεράστιο εμπόδιο όταν πρέπει να υιοθετηθεί ως πρακτική από τους ιατρούς και το προσωπικό των νοσοκομείων, οι οποίοι αφενός έχουν συνηθίσει, αφετέρου αναγνωρίζουν την ευκολία που τους προσφέρει η χρήση χειρογράφων σε πολλές περιπτώσεις. Για το λόγο αυτό παρατηρείται οργανωμένη προσπάθεια των διαχειριστών και διοικούντων των μονάδων υγείας ώστε να καθιερωθεί η ηλεκτρονική καταγραφή.

Η εταιρία **IBM**, επιχειρεί να αντιμετωπίσει το πρόβλημα της χρήσης διαφορετικών συνόλων δεδομένων. Χαρακτηριστικό παράδειγμα, το οποίο θα μελετηθεί εκτενέστερα στη συνέχεια, είναι ο **Watson**, ένας υπερυπολογιστής που διαθέτει μοναδικές ικανότητες χειρισμού φυσικής γλώσσας (**Natural Language Processing** ή **NLP**).

Επίσης η **Health Fidelity**, μία εταιρία που προσφέρει λύσεις στον τομέα της υγείας, κάνει χρήση εργαλείων επεξεργασίας φυσικής γλώσσας, με στόχο τη μετατροπή μη-δομημένων δεδομένων σε δομημένα.

Η μεγαλύτερη προοπτική των Big Data στον τομέα της υγείας, έγκειται στο συνδυασμό παραδοσιακών πηγών δεδομένων με νέες μορφές, οι οποίες θα πρέπει, ιδανικά, να παρέχουν πληροφορίες τόσο σε ατομικό όσο και σε πληθυσμιακό επίπεδο. Ήδη παρατηρούνται παραδείγματα όπου η χρήση διαφορετικών πηγών δεδομένων υπηρετούν την έρευνα με ταχύτερα και πιο αξιόπιστα αποτελέσματα. Εάν, για παράδειγμα, οι φαρμακευτικές εταιρίες, είχαν τη δυνατότητα να συγκεντρώσουν δημογραφικά κλινικά δεδομένα, θα μπορούσαν να διαθέσουν τα κατάλληλα φάρμακα πιο σύντομα στην αγορά, και κυρίως θα προσέγγιζαν το δομικό σκοπό τους, που είναι η παροχή του κατάλληλου φαρμάκου και την κατάλληλη στιγμή στον ασθενή.

Ταχύτητα (Velocity)

Τα περισσότερα δεδομένα στον τομέα της υγείας προέρχονται παραδοσιακά από στατικές πηγές, όπως ακτινογραφίες, έγγραφα νοσοκομείων, δελτία ασθενών, βιβλιάρια υγείας κ.ό.κ. Σε κάποιες εφαρμογές, όμως, κρίνεται απαραίτητη η επεξεργασία και αξιοποίηση των δεδομένων σε πραγματικό χρόνο, όπως για παράδειγμα η επίβλεψη της αρτηριακής πίεσης και της καρδιακής λειτουργίας σε κάποια εγχείρηση. Υπάρχουν περιπτώσεις επίσης που απαιτείται επεξεργασία των δεδομένων σε πιο αργούς σχετικά ρυθμούς, όπως ο προσδιορισμός των επιπέδων της γλυκόζης διαβητικών ατόμων σε καθημερινή βάση. Ένα ακόμα παράδειγμα αποτελούν οι πληροφορίες που αφορούν κάποια γνωστή ασθένεια, οι οποίες αυξάνονται ποσοστιαία με πολύ μικρότερο ρυθμό εν συγκρίσει με μία νέα επιδημία η οποία είναι σε έξαρση. Στη δεύτερη περίπτωση, τα δεδομένα καταφθάνουν με μεγάλο ρυθμό και συνιστούν “νέα” πληροφορία, ενώ επιτακτική είναι η ανάγκη άμεσης επεξεργασίας τους με σκοπό την αντιμετώπιση του προβλήματος σε πραγματικό χρόνο.

Οι μελλοντικές εφαρμογές της υγείας σε μονάδες εντατικής θεραπείας (**Intensive Care Units** ή **ICU**), όπως ο όσο το δυνατόν ταχύτερος εντοπισμός μολύνσεων, θα επιτρέπουν διαφορετική και πιο έγκαιρη και έγκυρη αντιμετώπιση, μειώνοντας έτσι κατά τεράστιο βαθμό τις περιπτώσεις θανάτων που παρατηρούνται από μολύνσεις σε αυτό το περιβάλλον. Όμοια, θα είναι δυνατή η αποτελεσματικότερη εποπτεία των νεογνών στους θαλάμους.

Οι εταιρίες που ειδικεύονται στην κατασκευή ιατρικών συσκευών , όπως η Baxter International, η Boston Scientific Corporation, η Hospira Inc. και η Medtronic Inc. έχουν κατασκευάσει συσκευές που επιτρέπουν την ανάλυση δεδομένων σε πραγματικό χρόνο σε ασθενοφόρα, σε χειρουργικά δωμάτια, σε νοσοκομεία, ακόμα και σε οικιακούς χώρους. Είναι γεγονός πως η δημιουργία συσκευών ικανών για επεξεργασία μεγάλου όγκου δεδομένων που καταφθάνουν ταχύτατα θα μπορούσε να σηματοδοτήσει την επανάσταση στην Ιατρική επιστήμη.

Εγκυρότητα (Veracity)

Ιδιαίτερα στον κλάδο της υγείας, η εγκυρότητα των δεδομένων είναι υψίστης σημασίας για δύο κυρίως λόγους: Σε αυτή βασίζονται αποφάσεις που καθορίζουν την ανθρώπινη ύπαρξη. Επίσης τα μη-δομημένα δεδομένα, αν και είναι μεγάλης σημασίας, συχνά εμπεριέχουν σφάλματα και λάθη ή είναι αδύνατο να αξιοποιηθούν για διάφορους λόγους. Αντιπροσωπευτικό παράδειγμα είναι η δυσκολία ανάγνωσης φαρμακευτικών συνταγών από τους ιατρούς.

Η μελέτη της αξιοπιστίας των δεδομένων στην υγεία αντιμετωπίζει περίπου τα ίδια ζητήματα που εγείρονται σε οικονομικές δοσοληψίες: Εγκυρότητα των στοιχείων του ασθενούς, ορθώς συμπληρωμένα δεδομένα στα πεδία που αφορούν το νοσοκομείο ή την κλινική, την ασφάλιση του ασθενούς, σύνδεση με τραπεζικούς λογαριασμούς, καταγραφή χρηματικών ποσών πληρωμής, εκκρεμών οφειλών κ.ό.κ. Βέβαια, στον κλάδο της υγείας παρατηρούνται στοιχεία τα οποία δεν παρατηρούνται σε άλλους τομείς, όπως πληροφορίες σχετικές με τη διάγνωση, τη θεραπεία, τη χορήγηση φαρμάκων, την περίθαλψη και όποια άλλη πληροφορία κρίνεται απαραίτητο να καταγραφεί. Η εγκυρότητα των στοιχείων αυτών είναι, σε κάθε περίπτωση, εξίσου σημαντική με αυτή των προαναφερθέντων.

Η βελτίωση των υπηρεσιών που αφορούν τη φροντίδα των ασθενών προϋποθέτει την αποφυγή λαθών και δημιουργεί τη διαρκή ανάγκη εύρεσης και αξιοποίησης των υψηλής ποιότητας δεδομένων, καθώς με αυτόν τον τρόπο η πρόοδος στην ασφάλεια των φαρμάκων και στην αποτελεσματικότητά τους, η μεγαλύτερης ακρίβειας διάγνωση, η καλύτερη στόχευση αντιμετώπισης της ασθένειας και εν τέλει η θεραπεία των ασθενών, πραγματοποιούνται σε άλλη βάση. Ωστόσο, δεδομένα τα οποία χαρακτηρίζονται από μεγάλη ποικιλία και ταχύτητα, απαιτούν μεγαλύτερο χρόνο ελέγχου εγκυρότητας, πριν αναλυθούν και αξιοποιηθούν, εγείροντας έτσι ζητήματα αξιοπιστίας των δεδομένων.

Ιδιαίτερα σημαντική είναι η διαπίστωση ότι τα ζητήματα ασφαλείας των υπό διάθεση φαρμάκων δεν είναι πάντα δυνατόν να επιλυθούν εάν οι δοκιμές πραγματοποιηθούν σε μικρό πλήθος ατόμων. Ουσιαστικά για τον αποτελεσματικό τους έλεγχο, απαιτούνται δοκιμές σε επαρκή πληθυσμό, που να αποτελεί αντιπροσωπευτικό δείγμα του συνόλου των ασθενών. Συνεπώς, γίνεται άμεσα αντιληπτό ότι ο εντοπισμός παθογενειών που συναντώνται σπάνια και υπό προϋποθέσεις είναι εφικτός μόνο μέσω μεγάλων συνόλων δεδομένων. Επίσης, κατά κοινή ομολογία, η χορήγηση κάποιου φαρμάκου δεν εγγυάται συγκεκριμένο αποτέλεσμα, καθώς πάντοτε θα υπάρχει μέρος του πληθυσμού που δεν

θα καλύπτει. Ακριβώς όμως επειδή οι περιπτώσεις αυτές είναι σπάνιες, η εγκυρότητα των δεδομένων είναι απαραίτητη ώστε να προσδιορισθούν οι σχέσεις και τα κοινά χαρακτηριστικά των περιπτώσεων όπου η θεραπεία αποτυγχάνει, ώστε να βελτιωθούν τα μοντέλα θεραπείας και χορήγησης φαρμάκων. Σε περίπτωση που σημαντικό ποσοστό των δεδομένων είναι χαμηλής ποιότητας η δυνατότητα αυτή εκλείπει.

Παρά τη σημασία που έχει η αξιοπιστία και εγκυρότητα των δεδομένων, είναι γεγονός πως μεγαλύτερη έμφαση δίνεται από τις εταιρίες που δραστηριοποιούνται στην αξιοποίηση των Big Data για τη βελτίωση άλλων χαρακτηριστικών, όπως η ταχύτερη επεξεργασία και ανάλυση των δεδομένων, παρά ο προσδιορισμός και περιορισμός λαθών. Εταιρίες όπως η IBM και η Google, καθώς και άλλοι οργανισμοί υγείας, όπως η NextBio, η Appistry, η Explorys Inc. και η Humedica έχουν εργαστεί σε μεγαλύτερο βαθμό στη μελέτη της εγκυρότητας των Big Data.

Αξία (Value)

Το κόστος της υγειονομικής περίθαλψης είναι μη βιώσιμο και εμφανίζει αύξηση σχεδόν σε όλα τα κράτη. Χαρακτηριστικά, εν έτει 2017, το ποσοστό του Ακαθάριστου Εγχώριου Προϊόντος (ΑΕΠ) των ΗΠΑ που δαπανάται για την υγειονομική περίθαλψη (περίπου 17%) θεωρείται ιδιαίτερα υψηλό. Επίσης οι περισσότερες εκτιμήσεις των μελλοντικών δαπανών υγειονομικής περίθαλψης δείχνουν αύξηση, δυσανάλογη σε σχέση με την οικονομία γενικότερα, σε όλες τις εθνικές οικονομίες.

Αυτό μπορεί να οφείλεται σε διάφορους παράγοντες, όπως:

- **Δημογραφικά στοιχεία:** Γήρανση του πληθυσμού, αύξηση προσδόκιμου χρόνου ζωής, περισσότερες χρόνιες παθήσεις, πιο δαπανηρή φροντίδα στα τελευταία στάδια του βίου.
- **Τεχνολογία:** Η προηγμένη και υψηλής τεχνολογίας Ιατρική απαιτεί συνήθως περισσότερα χρήματα σε σχέση τα επιτεύγματα χαμηλής τεχνολογίας της Ιατρικής του 20ού αιώνα.¹³
- **Ποιότητα των υπηρεσιών υγείας:** Ασυντόνιστες μονάδες υγείας, αναποτελεσματικές ροές τεχνολογίας, ιατρικά λάθη, νοσοκομειακές λοιμώξεις, ελλείψεις προσωπικού.

- **Διαρθρωτικά ζητήματα:** Θεσμικές ανεπάρκειες, κρούσματα διαφθοράς και απάτης, απόβλητα, στρεβλώσεις της αγοράς (σχέσεις εξάρτησης με κατασκευαστές, πληρωτές).

Ο ρυθμός αύξησης των δαπανών υγειονομικής περίθαλψης των ΗΠΑ, ανέρχεται σε ετήσια βάση περίπου στο 5% κατά την τελευταία δεκαετία, δεν είναι βιώσιμο και συνεισφέρει σημαντικά στα υψηλά εθνικά επίπεδα χρέους που προβλέπονται για τις επόμενες δύο δεκαετίες.¹⁴

Σύμφωνα με μία έκθεση αποτίμησης της υγείας που πραγματοποιήθηκε από την εταιρία Kauffman με θέμα «Βελτίωση της παραγωγικότητας και της ποιότητας της υγείας», επισημαίνεται ότι ο συνδυασμός ανεπαρκών πληροφοριών, τα ελλιπή κίνητρα για τον έλεγχο του κόστους και η αναποτελεσματικότητα στον τομέα της υγείας οδηγούν σε απίστευτη σπατάλη.¹⁵

Εκτιμάται ότι από τα 2.5 τρισεκατομμύρια δολάρια που δαπανήθηκαν για την υγειονομική περίθαλψη στις ΗΠΑ το 2010, τα 700 δισεκατομμύρια, δηλαδή το 28%, ήταν περιττά και αποτέλεσμα κακοδιαχείρισης.¹⁶

Πέραν όμως της μείωσης και του εξορθολογισμού των δαπανών, η χρήση και αξιοποίηση των Big Data στον τομέα της υγείας είναι δυνατόν να προσφέρει πολλαπλά οφέλη, όπως:¹²

- Δυνατότητα παροχής υπηρεσιών φροντίδας ειδικά σχεδιασμένων για τον κάθε ασθενή
- Αποτελεσματικότερος και ταχύτερος σχεδιασμός νέων φαρμάκων και κλινικών δοκιμών
- Καλύτερη πρόληψη στους πληθυσμούς υψηλού κινδύνου
- Άμεση πρόσβαση στη γνώση από ασθενείς και επαγγελματίες υγείας
- Συνδυασμός δεδομένων ετερογενών πηγών για ενιαία αποτελέσματα
- Μείωση κόστους μέσω εντοπισμού των μεθόδων που απαιτούν τις υψηλότερες δαπάνες
- Έλεγχος των πόρων που καταναλώνονται
- Άμεση σύγκριση αποτελεσμάτων διαφορετικών μεθόδων θεραπείας και πρόληψης
- Δυνατότητα Ιατρικής βασισμένης σε γεγονότα (evidence-based medicine)

- Σύγκριση αποτελεσματικότητας επαγγελματιών και φορέων υγείας
- Καλύτερη διαχείριση και κατανομή του ανθρωπίνου δυναμικού

2.2 Πηγές δεδομένων στην υγεία και μεθοδολογία

2.2.1. Πηγές δεδομένων στον τομέα της υγείας

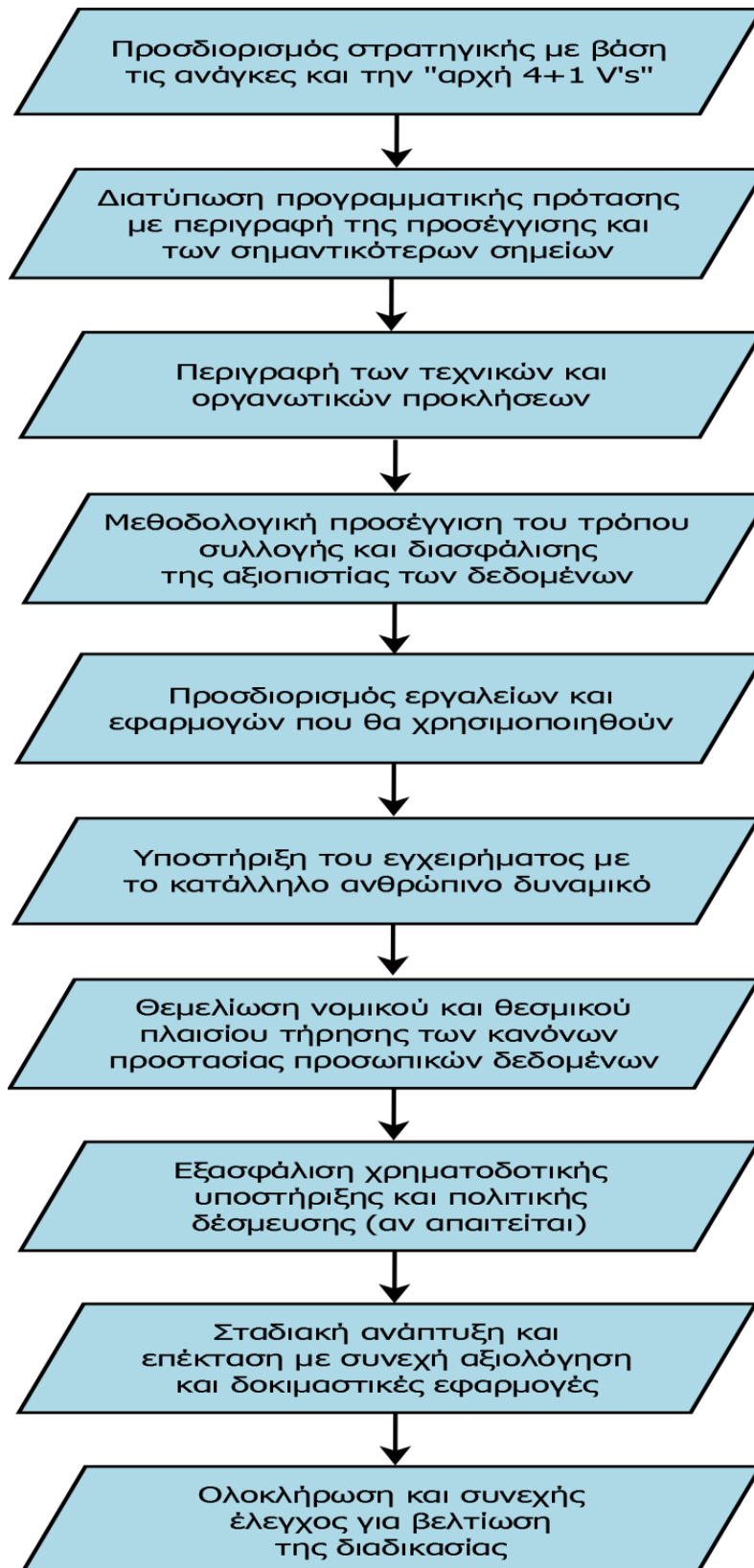
Στο σημείο αυτό έχουν παρουσιαστεί τα βασικά χαρακτηριστικά των δεδομένων που αφορούν τον τομέα της υγειονομικής περίθαλψης. Το γεγονός ότι είθισται ο χαρακτηρισμός Big Data να αποδίδεται σε συλλογές δεδομένων διαφορετικών πηγών επιβεβαιώνεται και στον υγειονομικό κλάδο, κάτι που γίνεται αντιληπτό μέσω της παρουσίας των πηγών, οι κυριότερες από τις οποίες είναι:¹

- Συστήματα κλινικών πληροφοριών
- Δεδομένα ασφαλιστικών οργανισμών
- Δημογραφικές και επιδημιολογικές εγγραφές
- Δεδομένα ιατρικών συσκευών και φαρμάκων
- Χρηματοδοτικές και οικονομικές συναλλαγές
- Δεδομένα βιοϊατρικών μετρήσεων
- Βάσεις γενετικών δεδομένων
- Αναζητήσεις διαδικτύου
- Μέσα κοινωνικής δικτύωσης
- Μη-δομημένα δεδομένα σημειώσεων και αλληλογραφίας

2.2.2 Προτεινόμενη μεθοδολογία για την ανάπτυξη αναλυτικής Big Data στην υγεία¹⁶

Παρά τη μεγάλη σημασία που έχει η κατανόηση των χαρακτηριστικών των Big Data και η πληροφόρηση σχετικά με τις διαθέσιμες πηγές, δεν αποτελεί μία απλη διαδικασία για ένα οργανισμό ή φορέα υγείας η στροφή προς την αξιοποίησή τους, καθώς υπάρχουν πολλοί παράγοντες που πρέπει να ληφθούν υπόψη. Η προτεινόμενη μεθοδολογία στον τομέα της υγείας προβλέπει τις ακόλουθες επιμέρους κινήσεις:

- Προσδιορισμός στρατηγικής με βάση τις ανάγκες και την αρχή «4+1V's».
- Διατύπωση προγραμματικής πρότασης με περιγραφή της προσέγγισης και των σημαντικότερων σημείων.
- Περιγραφή των τεχνικών και οργανωτικών προκλήσεων.
- Μεθοδολογική προσέγγιση του τρόπου συλλογής και διασφάλισης της αξιοπιστίας των δεδομένων.
- Προσδιορισμός εργαλείων και εφαρμογών που θα χρησιμοποιηθούν.
- Υποστήριξη του εγχειρήματος με το κατάλληλο ανθρώπινο δυναμικό.
- Θεμελίωση νομικού και θεσμικού πλαισίου τήρησης των κανόνων προστασίας των προσωπικών δεδομένων.
- Εξασφάλιση χρηματοδοτικής υποστήριξης και πολιτικής δέσμευσης (αν απαιτείται).
- Σταδιακή ανάπτυξη και επέκταση με συνεχή αξιολόγηση και δοκιμαστικές εφαρμογές.
- Ολοκλήρωση και συνεχής έλεγχος για βελτιώσεις της διαδικασίας.



Εικόνα 2.3: Σχηματικό διάγραμμα προτεινόμενης μεθοδολογίας για την αξιοποίηση των Big Data στην υγεία

2.3 Πεδία Εφαρμογής

Προτού πραγματοποιηθεί εκτενής μελέτη των στοιχείων που επιβεβαιώνουν την αξία της χρήσης των Big Data στην υπηρεσία της υγείας, υπάρχουν κάποια βασικά πεδία στα οποία είναι απαραίτητο να γίνει κατανοητή η συμβολή τους, αποδίδοντας το γενικότερο πλαίσιο εφαρμογής τους:

2.3.1 Προτεραιοποίηση επιλογών

Η συλλογή, επεξεργασία και διαχείριση επιδημιολογικών δεδομένων, καθώς και η συσχέτισή τους στον τομέα της κλινικής επιδημιολογίας και της αποτίμησης των παραγόντων κινδύνου, δίνει πολύ σημαντικές πληροφορίες για τον εντοπισμό και την προτεραιοποίηση των υγειονομικών στόχων. Για το λόγο αυτό, η εφαρμογή αυτή έχει μεγάλη βαρύτητα σε περιπτώσεις χρόνιων νοσημάτων, που αφορούν περισσότερο από το 40% του ενήλικου πληθυσμού. Μάλιστα, οι χρόνιες ασθένειες δημιουργούν περίπου το 70% της συνολικής ζήτησης των υπηρεσιών υγείας.¹⁶

Όμοια, η προτεραιοποίηση των επιλογών στους υγειονομικούς στόχους είναι δυνατόν να πραγματοποιηθεί σε εθνική, περιφερειακή και τοπική βάση με συνέπεια την ανάγκη διάθεσης ανθρώπινων, τεχνολογικών και οικονομικών πόρων, καθώς επίσης και των αντίστοιχων προϋπολογισμών.

Ένα χαρακτηριστικό παράδειγμα είναι η συσχέτιση του φύλου, της ηλικίας, των παραγόντων κινδύνου και των θεραπευτικών επιλογών, η οποία δύναται να βελτιώσει τους όρους κλινικής διακυβέρνησης με την παρακολούθηση και μέτρηση της αποτελεσματικότητας των διαθέσιμων θεραπευτικών βάσει κρίσιμων παραμέτρων, αναλόγως την περίπτωση.

2.3.2 Αποτίμηση Τεχνολογίας Υγείας

Η χρήση σύγχρονων εργαλείων κλινικής και οικονομικής διαχείρισης προσφέρει πολλά οφέλη, όπως η συμβολή στη μέτρηση της αποτελεσματικότητας και της αποδοτικότητας των παρεμβάσεων σε συνθήκες πραγματικής κλινικής πρακτικής. Παράλληλα, παρέχεται η δυνατότητα συγκεντρωτικής συλλογής επιδημιολογικών, κλινικών, οικονομικών και διαχειριστικών δεδομένων η οποία μπορεί να συνδράμει στην παραγωγή πληροφορίας συσχέτισης μεταξύ ανθρωπίνων και οικονομικών πόρων, σχετικά με τα υγειονομικά αποτελέσματα και τις εκβάσεις υγείας (health outcomes). Συνεπώς, η

πρακτική **Αποτίμησης Τεχνολογίας Υγείας (Health Technology Assessment)** με την ενσωμάτωση στη λήψη αποφάσεων των ερευνητικών αποτελεσμάτων από τις σχέσεις κόστους-αποτελεσματικότητας είναι δυνατόν να παρουσιάσει σημαντική συμβολή στην αποδοτικότητα και κατανομή των σπάνιων πόρων, καθώς και στην ερευνητική και ακαδημαϊκή δραστηριότητα.

Με τα εργαλεία αυτά καθίστανται εφικτό να προσδιορισθούν:

- η σύνδεση των δαπανόμενων οικονομικών πόρων με την επίτευξη υγειονομικού αποτελέσματος,
- ο υπολογισμός της επίπτωσης σε επιμέρους τομείς κάποιας ασθένειας στην υγεία του πληθυσμού και στο αντίστοιχο κόστος του υγειονομικού συστήματος,
- η βελτίωση της αποδοτικότητας του συστήματος και η ορθολογικοποίηση της χρήσης των υγειονομικών και οικονομικών πόρων,
- η εφαρμογή τεχνικών αξιολόγησης των τεχνολογιών υγείας και η εξαγωγή δεικτών,
- η σύγκριση διάφορων πολιτικών στον τομέα της υγείας,
- οι προοπτικές διαφοροποίησης των παρόντων πολιτικών διοίκησης στην υγεία μέσω μοντέλων πρόβλεψης.

2.3.3 Δεδομένα πραγματικής ζωής

Η μαζική συλλογή πληροφοριών από φορείς και οργανισμούς υγείας, συγκροτεί μια βάση πραγματικών κλινικών δεδομένων (real world data), αντίστοιχη με αυτή του Ηλεκτρονικού Φακέλου Υγείας, που θα περιγραφεί στη συνέχεια (αριθμός μητρώων ασφάλισης, φύλο, ηλικία, λοιπά δημογραφικά χαρακτηριστικά, ομάδα διαγνωστικής κατηγορίας, διαδικασία στρατηγικής θεραπείας, λήψη φαρμάκων και άλλων κλινικών διαδικασιών), η σωστή διαχείριση της οποίας είναι ικανή να συμβάλλει αποφασιστικά στην κυβερνησιμότητα του συστήματος υγείας, στη βελτίωση των επιλογών και στην προσέλκυση επενδυτικών δραστηριοτήτων, ενώ παράλληλα δύναται να αναδείξει τα σημεία και τα κέντρα άριστης κλινικής πρακτικής.

2.3.4 Στοιχεία κατανάλωσης υπηρεσιών

Οι υπηρεσίες διαχείρισης ηλεκτρονικών βάσεων διευκολύνουν τη συλλογή ιατρικών δεδομένων, με απαραίτητη προϋπόθεση τον έλεγχο της ποιότητας. Συνεπώς, η συνεργασία μεταξύ φορέων υγείας, ιατρικών σχολών και άλλων επαγγελματιών που δραστηριοποιούνται στο χώρο της υγείας ενδέχεται να οδηγήσει σε δημιουργία σύγχρονων βάσεων δεδομένων ανά ασθένεια ή ομάδες ασθενειών. Η συνδρομή τέτοιων βάσεων δεδομένων προσφέρει επιπλέον όφελος στην προσέλευση σημαντικών επενδύσεων.

Η ανάλυση και η ταξινόμηση των δεδομένων σε μεγάλα τμήματα, και ειδικότερα σε μέρη τα οποία συσχετίζονται με τη χρήση αγαθών και υπηρεσιών υγείας, μπορεί να μειώσει την άγνοια των προμηθευτών υγείας (providers ignorance) όσων αναφέρονται σε φάρμακα και τεχνολογικά προϊόντα. Παρόλο που η συλλογή και ανάλυση δεδομένων πραγματικής κλινικής πρακτικής (real world data) αποτελεί σημαντική προοπτική, η ανάπτυξη μηχανισμών συλλογής δεδομένων (αρχαία ασθενών) συνιστούν χρονοβόρες και δαπανηρές διαδικασίες.

2.3.5 Κλινική έρευνα και έρευνα υπηρεσιών υγείας

Οι δυνατότητες των ερευνητών διευκολύνονται σε μεγάλο βαθμό από την ελεύθερη πρόσβαση επιδημιολογικών, διαχειριστικών και κλινικών δεδομένων του υγειονομικού τομέα, γεγονός που αναμένεται να συμβάλλει στην αύξηση του όγκου των δεδομένων και τη βελτίωση της ποιότητας της επιστημονικής έρευνας, αλλά και στην επιστημονική εμπέλεια των ιδρυμάτων και της ερευνητικής και ακαδημαϊκής κοινότητας.

Μάλιστα, η συμβολή στη μεταφορά των αποτελεσμάτων της έρευνας στην άσκηση κλινικής πρακτικής και στην εγκαθίδρυση μεθόδων τεκμηριωμένης Ιατρικής (evidence-based medicine), είναι επίσης μια σημαντική συμβολή στην προστιθέμενη αξία της Ιατρικής περίθαλψης.

2.3.6 Βελτίωση της κυβερνησιμότητας στον τομέα της υγείας

Η μετάβαση σε μια κατάσταση ηλεκτρονικής κυβερνησιμότητας (e-governing) δύναται να βοηθήσει στην υποστήριξη των πολιτών για έγκυρη και ορθή ενημέρωση, στην επέκταση των υπηρεσιών σε κομβικές περιοχές που ευνοούν τους χρήστες, στη διαφάνεια και στη διασφάλιση συνεχούς πρόσβασης σε ένα ενημερωμένο κοινό.¹⁷

ΚΕΦΑΛΑΙΟ 3

ΕΡΓΑΛΕΙΑ ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΓΙΑ ΤΗΝ ΑΞΙΟΠΟΙΗΣΗ ΤΩΝ BIG DATA

3.1 Βασικά εργαλεία

Εφόσον έχουν γίνει κατανοητά τα βασικά ζητήματα που αφορούν τη χρήση, συλλογή και διαχείριση των Big Data στον τομέα της υγείας, κρίνεται σκόπιμο να διερευνηθούν τα παρεχόμενα από την τεχνολογία εργαλεία αξιοποίησης των δεδομένων.

Όπως συμβαίνει σχεδόν πάντοτε σε τομείς χρήσης λογισμικού, υπάρχει και στη χρήση των Big Data η δυνατότητα επιλογής μεταξύ χρήσης ελεύθερου λογισμικού, αλλά και εμπορικών λύσεων, οι οποίες απαιτούν τη χρήση οικονομικών πόρων. Η πλατφόρμα που θα επιλεγθεί, σε κάθε περίπτωση, θα πρέπει να χειρίζεται την εισαγωγή, την επεξεργασία, την αποθήκευση και την αναζήτηση των δεδομένων, καθώς επίσης να παρέχει δυνατότητες ανάλυσής τους. Στο σημείο αυτό θα πραγματοποιηθεί μία παρουσίαση των βασικών επιλογών που παρέχονται.

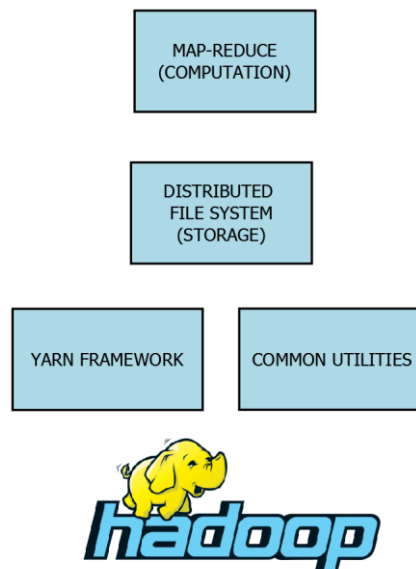
3.1.1 Apache Hadoop

Πρόκειται για ελεύθερου κώδικα λογισμικό (open-source software), το οποίο χρησιμοποιείται για κατανεμημένη αποθήκευση και επεξεργασία Big Data χρησιμοποιώντας το μοντέλο Map-Reduce.

Το Hadoop αποτελείται από τα εξής δομικά στοιχεία:

- Το **Hadoop Common Utilities** που περιέχει βασικές βιβλιοθήκες και λειτουργίες που απαιτούνται από τα υπόλοιπα στοιχεία.
- Το **Hadoop Distributed File System (HDFS)** που διαχειρίζεται την αποθήκευση κατανεμημένων δεδομένων.
- Το **Hadoop YARN Framework**, το οποίο αποτελεί μία πλατφόρμα διαχείρισης πόρων. Ουσιαστικά, είναι υπεύθυνο για τη διαχείριση των υπολογιστικών πόρων σε συστάδες και για τον προγραμματισμό των εφαρμογών των χρηστών.
- Το **Hadoop Map-Reduce** που αποτελεί υλοποίηση του μοντέλου Map-Reduce για κατανεμημένη επεξεργασία μεγάλης κλίμακας δεδομένων.

Δομή Apache Hadoop



Εικόνα 3.1: Η δομή του Apache Hadoop

Ένα υπολογιστικό σύστημα που εκτελεί την εφαρμογή Hadoop αποτελείται από υπολογιστικές συστάδες (**clusters**) οι οποίες απαρτίζονται από εμπορικό υλικό (**commodity hardware**). Η δομή του Hadoop βασίζεται στην υπόθεση ότι οι αστοχίες υλικού (**hardware failures**), δηλαδή οι δυσλειτουργίες στα ηλεκτρονικά στοιχεία των υπολογιστικών συστημάτων - είναι συχνές κατά τη διαχείριση μεγάλου όγκου δεδομένων και οφείλει η ίδια η εφαρμογή να τις διαχειρίζεται αποδοτικά.¹⁸

Ο πυρήνας του Hadoop αποτελείται από ένα τμήμα αποθήκευσης, γνωστό ως Hadoop Distributed File System (HDFS) και ένα τμήμα επεξεργασίας, που βασίζεται στο μοντέλο Map-Reduce που αναφέρθηκε προηγουμένως. Το Hadoop χωρίζει τα δεδομένα σε μεγάλα τμήματα (**blocks**) και τα κατανέμει μεταξύ διαφόρων υπολογιστικών κόμβων που συνιστούν το υπολογιστικό σύστημα. Στη συνέχεια, μεταφέρει τον κώδικα που πρόκειται να εκτελεστεί στους κόμβους ώστε να πραγματοποιηθεί παράλληλη, δηλαδή ταυτόχρονη επεξεργασία των δεδομένων στους κόμβους αυτούς. Ουσιαστικά, διενεργείται αξιοποίηση της ιδιότητας της τοπικότητας των δεδομένων (**data locality**) και οι κόμβοι διαχειρίζονται τα επιμέρους δεδομένα στα οποία έχουν πρόσβαση.¹⁹

Η χρήση της παραλληλίας, αποτελεί κλασική προσέγγιση βελτίωσης της αποδοτικότητας εφαρμογών λογισμικού. Μάλιστα αποδεδειγμένα, η χρήση της τοπικότητας των δεδομένων κατ' αυτό τον τρόπο από εμπορικά συστήματα, παρέχει καλύτερα αποτελέσματα από αυτά που

προσφέρουν εξελιγμένοι υπερυπολογιστές (supercomputers), οι οποίοι βασίζονται σε παράλληλα συστήματα αρχείων (parallel file systems), όπου ο υπολογισμός και τα δεδομένα διαμοιράζονται μέσω υψηλής ταχύτητας δικτύου.²⁰ Το μοντέλο Map-Reduce είναι ένα μοντέλο προγραμματισμού αποτελούμενο από μια συνάρτηση απεικόνισης (Map) η οποία φιλτράρει και διατάσσει τα δεδομένα στους κόμβους και μία συνάρτηση μείωσης (Reduce), η οποία πραγματοποιεί υπολογισμούς καταμέτρησης στους υπολογιστικούς κόμβους.

Κλασικό παράδειγμα χρήσης του μοντέλου Map-Reduce αποτελεί ένα πρόγραμμα μέτρησης λέξεων (word-count) σε ένα σύνολο αρχείων. Στην περίπτωση αυτή, η συνάρτηση Map απεικονίζει το σύνολο των αρχείων τμηματικά σε κάθε υπολογιστικό κόμβο και η συνάρτηση Reduce μετράει το πλήθος των λέξεων σε κάθε υπολογιστικό κόμβο, λαμβάνοντάς τελικά το ζητούμενο αθροιστικά.

Τέλος, αξίζει να σημειωθεί πως αν και η βασική δομή του Hadoop συνίσταται από τα στοιχεία που ήδη αναφέρθηκαν, συχνά χρησιμοποιούνται επεκτάσεις από την Apache που εμπλουτίζουν τις δυνατότητες του Hadoop, αναλόγως την περίπτωση, οι σημαντικότερες από τις οποίες είναι: Apache HBase, Apache Pig, Apache Hive, Apache Phoenix, Apache Spark, Apache ZooKeeper, Apache Flume, Apache Sqoop, Apache Storm.²¹

3.1.2 Apache Spark

Αποτελεί επίσης ελεύθερου κώδικα λογισμικό για επεξεργασία Big Data. Δημιουργήθηκε αρχικά στο Πανεπιστήμιο Berkeley, της California και στη συνέχεια παραχωρήθηκε αφιλοκερδώς στην Apache Software Foundation. Δημιουργήθηκε μετά το Hadoop και ουσιαστικά προσφέρει κάποια πλεονεκτήματα, τα οποία θα αναλυθούν στη συνέχεια. Το Spark προσφέρει στον προγραμματιστή μία διεπαφή (Interface) επικεντρωμένη σε μία δομή δεδομένων, γνωστή ως **Ελαστικό Κατανεμημένο Σύνολο Δεδομένων (Resilient Distributed Dataset ή RDD)** και πρόκειται για μια συλλογή κατανεμημένων αντικειμένων σε ένα σύνολο υπολογιστικών κόμβων η οποία διασφαλίζει αποτελεσματική διαχείριση αστοχιών υλικού, όπως ακριβώς το Hadoop.²²

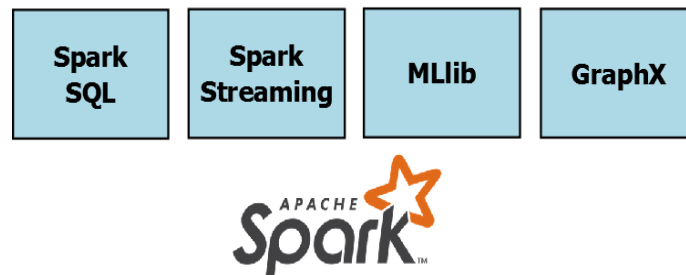
Ο λόγος δημιουργίας του Spark είναι ορισμένοι δομικοί περιορισμοί που επιβάλλονται από το μοντέλο Map-Reduce του Hadoop. Συγκεκριμένα, απαιτείται γραμμική ροή δεδομένων ως είσοδος από το δίσκο σε κατανεμημένα συστήματα, κατάλληλη επεξεργασία σύμφωνα με τις συναρτήσεις Map

και Reduce και τέλος γραμμικού χρόνου αποθήκευση των δεδομένων στο δίσκο. Αντίθετα, το Spark παρέχει τη δυνατότητα πραγματοποίησης των υπολογισμών σε διαμοιραζόμενη μνήμη, όπου η ταχύτητα είναι σημαντικά μεγαλύτερη από την αντίστοιχη σε δίσκο. Με τον τρόπο αυτό, καθίσταται δυνατή η εφαρμογή επαναληπτικών αλγορίθμων που πραγματοποιούν πολλαπλές φορές πρόσβαση στα δεδομένα σε κάθε επανάληψη, χωρίς αυτό να συμβαίνει εις βάρος του χρόνου υπολογισμού, αφού ο χρόνος πρόσβασης σε δεδομένα μνήμης είναι ταχύτερος και «πλησιέστερα» στον επεξεργαστή των υπολογιστικών κόμβων.²³ Σύμφωνα μάλιστα με πληροφορίες που βρίσκονται στην επίσημη ιστοσελίδα του Spark, είναι δυνατόν να εκτελεστούν εφαρμογές έως και 100 φορές ταχύτερα στη μνήμη και έως 10 φορές ταχύτερα στο δίσκο, εν συγκρίσει με το Hadoop. Είναι δυνατόν μάλιστα να εκτελεστεί πάνω στον πυρήνα του Hadoop.

Το Apache Spark πέραν του τρόπου διαχείρισης των Big Data, προσφέρει τις εξής βασικές επεκτάσεις:

- **Spark SQL:** Επιτρέπει ερωτήματα (queries) σε δεδομένα με χρήση SQL, σε συνδυασμό με τις γλώσσες προγραμματισμού Java, Scala, Python και R.
- **Spark Streaming:** Καθιστά εφικτή την επεξεργασία δεδομένων σε ροή, δηλαδή δεδομένων που εισέρχονται στο σύστημα ενώ βρίσκονται ήδη σε εξέλιξη υπολογισμοί στα προηγούμενα δεδομένα. Αυτό το χαρακτηριστικό είναι πολύ σημαντικό, καθώς στο Hadoop δεν μπορούν να προστίθενται νέα δεδομένα κατά τη διάρκεια της επεξεργασίας, αλλά πρέπει να είναι διαθέσιμο όλο το σύνολό τους όταν εκκινεί μία Map-Reduce διαδικασία. Υποστηρίζονται οι γλώσσες προγραμματισμού Java, Scala και Python.
- **MLlib:** Πρόκειται για μία βιβλιοθήκη μηχανικής μάθησης (**Machine Learning Library**) η οποία δίνει τη δυνατότητα εκτέλεσης αλγορίθμων αυτού του είδους έως και 100 φορές ταχύτερα από το Hadoop.
- **GraphX:** Παρέχει ένα **API (Application Programming Interface)** για τα δεδομένα σε μορφή γραφημάτων, επιτρέποντας μάλιστα υπολογισμούς με χρήση επαναληπτικών αλγορίθμων με αποδοτικό τρόπο.

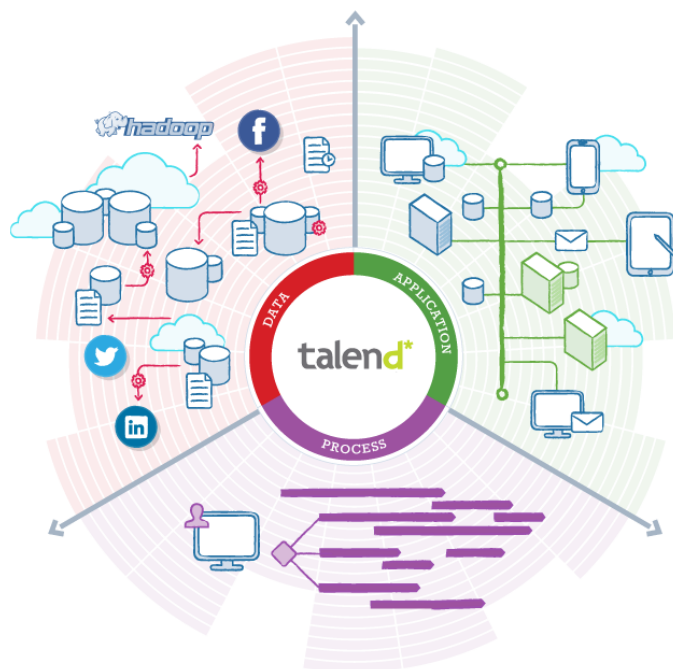
Δομή Apache Spark



Εικόνα 3.2: Η δομή του Apache Spark

3.1.3 Talend

Το Talend είναι μια πλατφόρμα ανοικτού κώδικα, βασισμένη στο μοντέλο Hadoop που προσφέρει μια σειρά προϊόντων διαχείρισης Big Data. Το βασικότερο στοιχείο είναι το Master Data Management (MDM), το οποίο έχει τη δυνατότητα να επεξεργάζεται δεδομένα σε πραγματικό χρόνο, να αξιοποιεί άλλες εφαρμογές, να ενσωματώνει τα δεδομένα τους και να εκτελεί διάφορες διαδικασίες, όπως εκτιμήσεις της ποιότητας των Big Data. Παρέχεται δωρεάν και προσφέρει αρκετές δυνατότητες, καθιστώντας το καλή επιλογή για πολλές ανάγκες.²⁴



Εικόνα 3.3: Η διαχείριση των δεδομένων από το Talend βάσει του μοντέλου Hadoop

3.2 Επιπρόσθετα εργαλεία

3.2.1 Χρήση NoSQL Databases

Οι 10Gen, Cloudera και Amazon ήταν οι πρώτες εταιρίες που διαμόρφωσαν πλατφόρμες εκμετάλλευσης Big Data με δυνατότητα υποστήριξης του Apache Hadoop και τεχνολογιών για μη σχεσιακές βάσεις δεδομένων (**NoSQL Databases**), ενώ στην πορεία εμφανίστηκαν και άλλες, όπως οι Amazon, DataStax, Neo Technologies, Hortonworks, Platfora, 10Gen και CouchBase.

Οι σχεσιακές βάσεις δεδομένων, δομημένες βάσει της γλώσσας SQL, ήταν για πολλά χρόνια ο πιο δημοφιλής τρόπος διαχείρισης δεδομένων για οργανισμούς και επαγγελματίες στον τομέα της τεχνολογίας. Με την έλευση των Big Data, τα οποία χαρακτηρίζονται τόσο από το μεγάλο μέγεθος, όσο και από την ποικιλομορφία στη δομή τους, καθίσταται απαραίτητη η δυνατότητα επεξεργασίας δεδομένων σε μεγάλη κλίμακα με σκοπό την εξαγωγή ενιαίων συμπερασμάτων. Στο πρόβλημα της διαχείρισης των δεδομένων αυτών, τα συστήματα που βασίζονται στην SQL δεν είναι δυνατόν να προσφέρουν αυτόνομα τη λύση.

Το πρόβλημα αντιμετωπίζεται μέσω της χρήσης NoSQL βάσεων δεδομένων (NoSQL databases), οι οποίες παρέχουν δυνατότητες δυναμικής διαχείρισης δεδομένων (dynamic data management), αυξημένης ευελιξίας (flexibility) και κλιμάκωσης (scalability) εν συγκρίσει με τις σχεσιακές βάσεις δεδομένων. Τα χαρακτηριστικά τους, τις καθιστούν ιδανικές για διαχείριση μεγάλου μεγέθους, μη-ομοιογενών δεδομένων τα οποία ανανεώνονται συχνά και πολλές φορές μεταβάλλονται οι τύποι (formats) των πεδίων (fields) των δεδομένων, πέραν των ίδιων των δεδομένων. Η διαδικασία αυτή είναι εξαιρετικά χρονοβόρα και σε πολλές περιπτώσεις ανέφικτη σε σχεσιακά συστήματα διαχείρισης βάσεων δεδομένων.

Στο σημείο αυτό είναι απαραίτητο να διευκρινιστεί το γεγονός πως παρά τα οφέλη που παρέχουν οι NoSQL βάσεις δεδομένων, δεν είθισται να χρησιμοποιούνται αυτόνομα για τη διαχείριση Big Data. Όπως αναφέρει ο επιστήμονας μεγάλων δεδομένων (Big Data Scientist) William McKnight, «οι NoSQL βάσεις δεδομένων πρέπει να χρησιμοποιούνται μόνο όταν δεν είναι δυνατή η χρήση σχεσιακών βάσεων», ώστε να βελτιστοποιείται η αξιοποίηση των δεδομένων.

Οι βασικές επιλογές NoSQL βάσεων δεδομένων, είναι οι **MongoDB**, **Neo4j**, **CouchBase**, **DynamoDB**, **HBase** και **Cassandra**. Στις εταιρίες που δραστηριοποιούνται στο χώρο της υγείας, έχει κυριαρχήσει η χρήση της MongoDB, έναντι των υπολοίπων, οι οποίες πάντως χρησιμοποιούνται σε αρκετές εφαρμογές.

Η MongoDB παρέχεται από την εταιρία 10Gen και μπορεί να συνδυαστεί αποδοτικά με τη χρήση τεχνολογιών JSON (Java Script Object Notation), XML, κ.ό.κ. Σύμφωνα με τα όσα υποστηρίζει η ίδια η εταιρία, η MongoDB είναι ευέλικτη, εύκολη στη χρήση, παρέχει υψηλές αποδόσεις, διαθεσιμότητα και αυτόματη κλιμάκωση.²⁵ Μεταξύ άλλων σημαντικών χαρακτηριστικών, διαθέτει τη δυνατότητα αναζήτησης σε κείμενα και σύνδεσης με το Hadoop.

Μερικά ενδεικτικά παραδείγματα λύσεων που παρέχει η MongoDB στον τομέα της υγείας είναι:

- **Δημιουργία πλήρους προφίλ ασθενούς** αποτελούμενο από όλες τις εξετάσεις που του έχουν διενεργηθεί και εξαγωγή χρήσιμων σχέσεων μεταξύ αυτών μέσω τεχνικών εξόρυξης δεδομένων (**data mining**). Είναι εύκολη η τροποποίηση και η προσθήκη νέων δεδομένων εξετάσεων στο προφίλ καθώς και η σύγκριση παλαιών και νέων δεδομένων.
- **Έγκαιρη διάγνωση σπάνιων ασθενειών**. Γίνεται εφικτό να εντοπιστούν σπάνιες ασθένειες που μπορεί να έχουν ένα κοινό σύνολο συμπτωμάτων, αλλά το κάθε ένα εξ αυτών ή κάποιο υποσύνολό τους δεν αποτελούν επίφοβη ένδειξη. Η παρατήρηση αυτή αποκτά μεγαλύτερη σημασία δοθέντος του γεγονότος ότι οι θεράποντες ιατροί πραγματοποιούν διαγνώσεις στηριζόμενοι κυρίως στην εμπειρία και στο ιστορικό των ασθενών που έχουν εξετάσει στο παρελθόν, καθιστώντας τη διαδικασία διάγνωσης σπάνιων ασθενειών σε πρώιμο στάδιο εξαιρετικά δύσκολη, λόγω της φύσεως της ανθρώπινης συλλογιστικής. Εφαρμογές που διαθέτουν μεγάλο σύνολο στατιστικών δεδομένων είναι πολύ εύκολο να εξάγουν δείκτες ταύτισης με ασθένειες και να προσφέρουν ένα εξαιρετικά χρήσιμο εργαλείο για το ιατρικό προσωπικό.
- **Περιορισμός εξάπλωσης επιδημιών σε πρώιμο στάδιο**. Η αξιοποίηση των Big Data έχει τη δυνατότητα να σώσει ανθρώπινες ζωές, σε περιπτώσεις που καμία άλλη μέθοδος δεν είναι εφικτό να αποδώσει. Η συλλογή δεδομένων για νέες ασθένειες – που εγκυμονούν κίνδυνο μαζικής εξάπλωσης - μπορεί να αξιοποιηθεί μέσω εφαρμογών οι οποίες λειτουργούν ως εργαλείο για το ιατρικό προσωπικό και επιτρέπουν την εξαγωγή δεικτών επικινδυνότητας, όπως ταχύτητα εξάπλω-

σης, πλήθος νοσούντων, συμπτώματα, σύγκριση με δεδομένα παλαιότερων επιδημιών και προσφέρουν τη δυνατότητα έγκαιρης εφαρμογής δημογραφικών μέτρων περιορισμού, όποτε κρίνεται απαραίτητο.

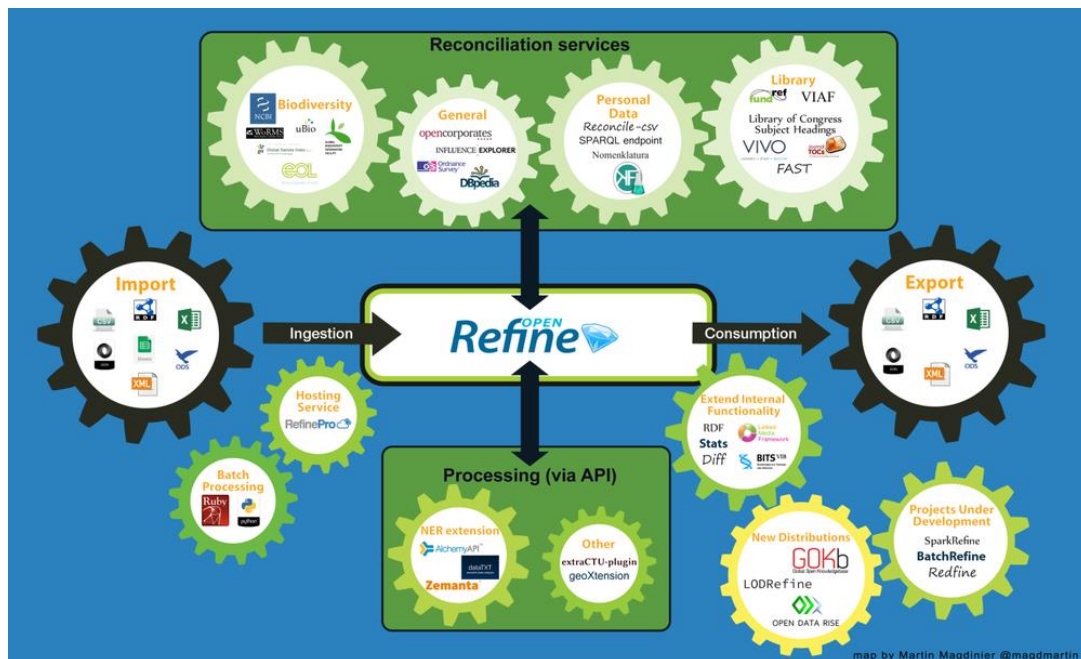
- **Άμεση γνωμάτευση σε πραγματικό χρόνο.** Σε περιπτώσεις εργαστηριακών δεδομένων από εξετάσεις που διενεργούνται σε ασθενείς, είναι δυνατή η άμεση εξαγωγή ποσοτικών συμπερασμάτων από το ιατρικό προσωπικό, καθώς παρέχεται η δυνατότητα οπτικής απεικόνισης μετρήσεων πάσης φύσεως και πηγής δεδομένων σε ενιαία διαγράμματα μέσω γραφημάτων, δίχως να απαιτείται ανεξάρτητη μελέτη των επιμέρους εξετάσεων από τον θεράποντα ιατρό.

3.2.2 Εργαλεία εκκαθάρισης των δεδομένων

Σε πολλές περιπτώσεις είναι χρήσιμο να πραγματοποιείται εκκαθάριση και μετατροπή των μη-δομημένων δεδομένων σε δομημένα, ιδιαίτερα όταν αυτά προέρχονται από πηγές του διαδικτύου και διαθέτουν μεγάλη μορφολογική ποικιλία.

OpenRefine

Το OpenRefine είναι ένα εργαλείο ελεύθερου κώδικα το οποίο χρησιμοποιείται για την εκκαθάριση μεγάλων συνόλων δεδομένων. Παρέχει διάφορες επιλογές εκκαθάρισης (Reconciliation Services) και μία προγραμματιστική διεπαφή (Application Programming Interface) για τη μετατροπή των δεδομένων εισόδου σε πιο χρήσιμη και ευκολότερα αξιοποιήσιμη μορφή.



Εικόνα 3.4: Τα επίπεδο λειτουργικότητας του OpenRefine

Data Cleaner

Το Data Cleaner χρησιμοποιείται για την εκκαθάριση των δεδομένων. Στηρίζεται στην παραδοχή ότι η διαχείριση των δεδομένων είναι χρονοβόρα και επίπονη διαδικασία.²⁶

Η λειτουργία του βασίζεται στα εξής ακολουθιακά βήματα:

- Εισαγωγή δεδομένων
- Συγχώνευση συνόλων δεδομένων
- Επανακατασκευή ελλιπών δεδομένων
- Προτυποποίηση
- Κανονικοποίηση
- Απαλοιφή διπλότυπων
- Επιβεβαίωση και Επέκταση
- Εξαγωγή δεδομένων



Εικόνα 3.5: Ακολουθιακό σχηματικό διάγραμμα του Data Cleaner

3.2.3 Εργαλεία Απεικόνισης

Η ποικιλία των Big Data σε συνδυασμό με τον τεράστιο όγκο τους, καθώς και όλες οι ιδιομορφίες που συνήθως τα συνοδεύουν, καταδεικνύουν την ανάγκη αποτελεσματικής απεικόνισής τους. Για το λόγο αυτό είναι ιδιαίτερα σημαντική η συνεισφορά των εργαλείων που παρέχουν αυτή τη δυνατότητα και κατατάσσονται στην κατηγορία Οπτικής Αναλυτικής (**Visual Analytics**). Μέσω αυτών τα Big Data γίνονται κατανοητά από τον ενδιαφερόμενο, δίχως να απαιτείται στις περισσότερες περιπτώσεις, τεχνική γνώση.

Ακολουθούν οι σημαντικότερες εφαρμογές που προσφέρονται:

Tableau

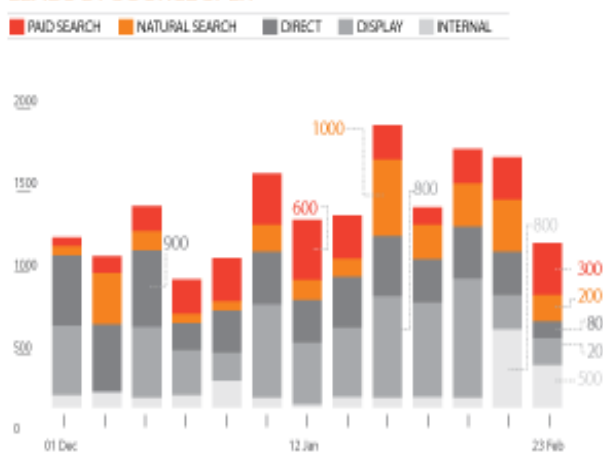
Το Tableau είναι ένα εργαλείο απεικόνισης δεδομένων με κύρια στόχευση την εύκολη δημιουργία διαγραμμάτων, χωρίς να είναι απαραίτητη η γνώση προγραμματισμού. Η μεγαλύτερη καινοτομία

έγκειται στη δυνατότητα αξιοποίησης δεδομένων που βρίσκονται στο διαδίκτυο, χωρίς να είναι απαραίτητη η λήψη τους, μέσω μίας διεπαφής που παρέχει η εφαρμογή. Γενικά, θεωρείται η πιο ευέλικτη και πλούσια σε δυνατότητες εφαρμογή για την απεικόνιση στατιστικών στοιχείων και Big Data.

Υπάρχουν πέντε εκδόσεις της εφαρμογής με διαφορετικές λειτουργίες και δυνατότητες υποστήριξης. Για τους αρχάριους προτείνεται η έκδοση Tableau Public, η οποία παρέχεται δωρεάν και ευνοεί την εξοικείωση με την εφαρμογή και τη χρήση της.

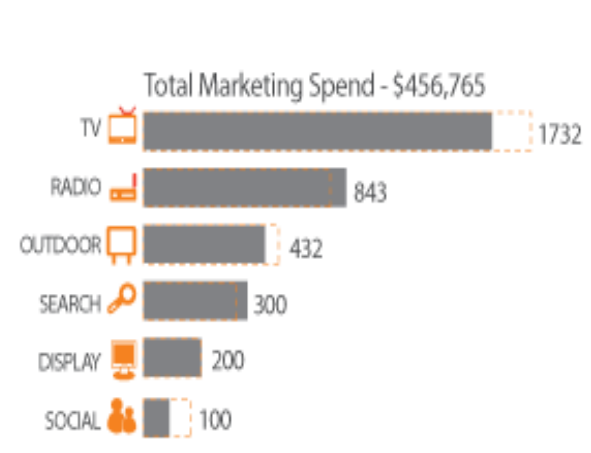
Στην ακόλουθη εικόνα παρουσιάζονται διάφορα γραφήματα που απεικονίζουν χρηματικά ποσά που δαπανήθηκαν για την έρευνα και διαφήμιση φαρμακευτικών προϊόντων, όπως φαίνεται με τρόπο ιδιαίτερα φιλικό προς το χρήστη.

LEADS BY SOURCE SPLIT



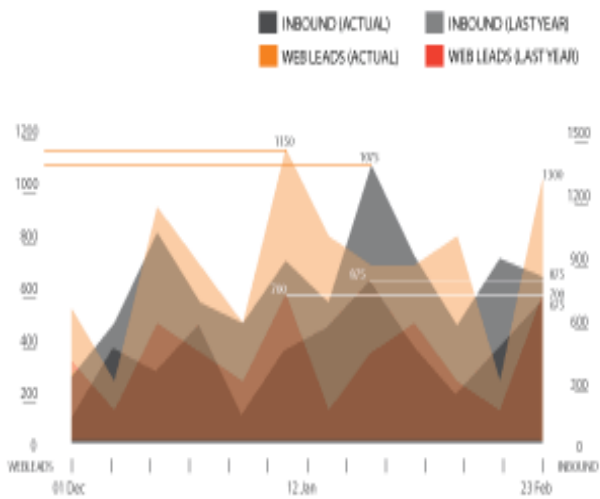
Εικόνα 3.6: Χρηματικά ποσά που διατίθενται για έρευνα

TOTAL SPEND



Εικόνα 3.7: Κονδύλια για διαφήμιση φαρμάκων ανά μέσο

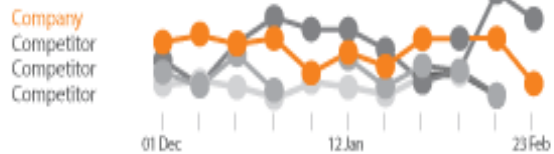
WEB LEADS AND INBOUND (ACTUAL AND LAST YEAR)



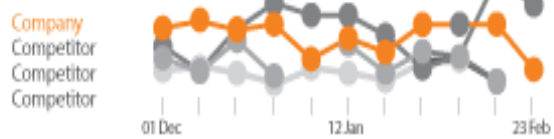
Εικόνα 3.8: Αναπαράσταση πολλών μεγεθών προς χρόνο(i)

COMPETITORS (SEO AND SEM)

SEO VISIBILITY SCORE V COMP

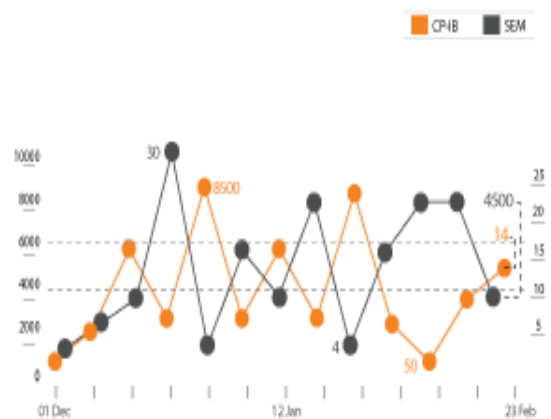


SEM COMP SPEND



Εικόνα 3.9: Αναπαράσταση πολλών μεγεθών προς χρόνο(ii)

COST PER INBOUND AND SEM



Εικόνα 3.10: Αναπαράσταση μεγεθών ανά περίοδο

MEDIA SPEND BY STATE

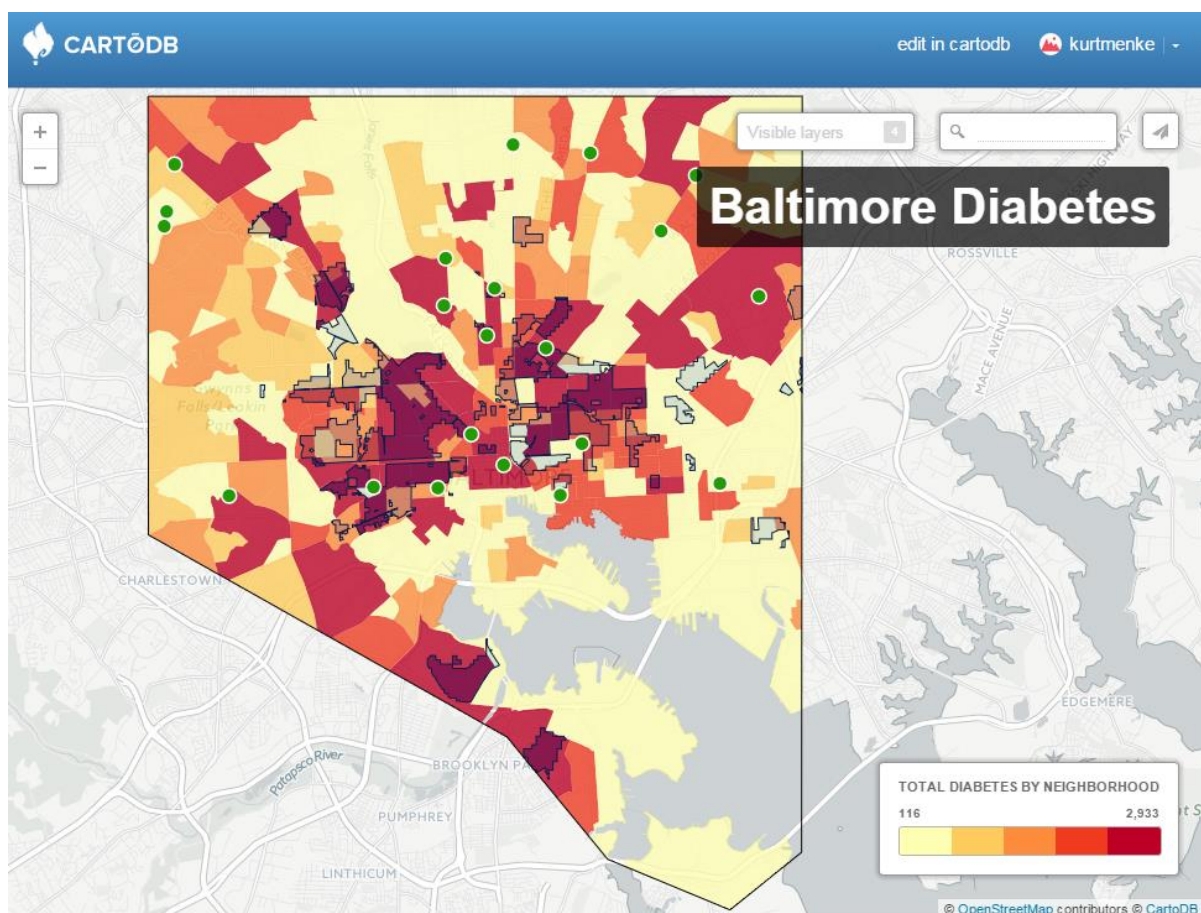


Εικόνα 3.11: Αναπαράσταση μεγεθών ανεξάρτητα με ενιαία μορφή

CartoDB

Δίχως να προσφέρει τόσες δυνατότητες γραφημάτων, όσες το Tableau, το CartoDB είναι μία εφαρμογή η οποία δημιουργεί χάρτες και χρησιμοποιείται κυρίως για αναπαράσταση πληροφορίας με κριτήριο την τοπικότητα των φαινομένων. Ενδείκνυται για μελέτη Big Data, καθώς διαχειρίζεται πολλών ειδών δεδομένα και τύπους αρχείων, παρέχει τη δυνατότητα ενιαίων αποτελεσμάτων και διαθέτει πρότυπα σύνολα δεδομένων με τα οποία είναι εύκολο να εξοικειωθεί κανείς.

Σύμφωνα με αυτά, το CartoDB , αποτελεί εξαιρετική επιλογή για χρήση σε χάρτες. Η ακόλουθη εικόνα αναπαριστά με διαφορετική χρωματική διαβάθμιση το πλήθος των διαβητικών ατόμων στην Πολιτεία της Βαλτιμόρης και προέρχεται από συγχώνευση διαφορετικών ετερογενών πηγών.

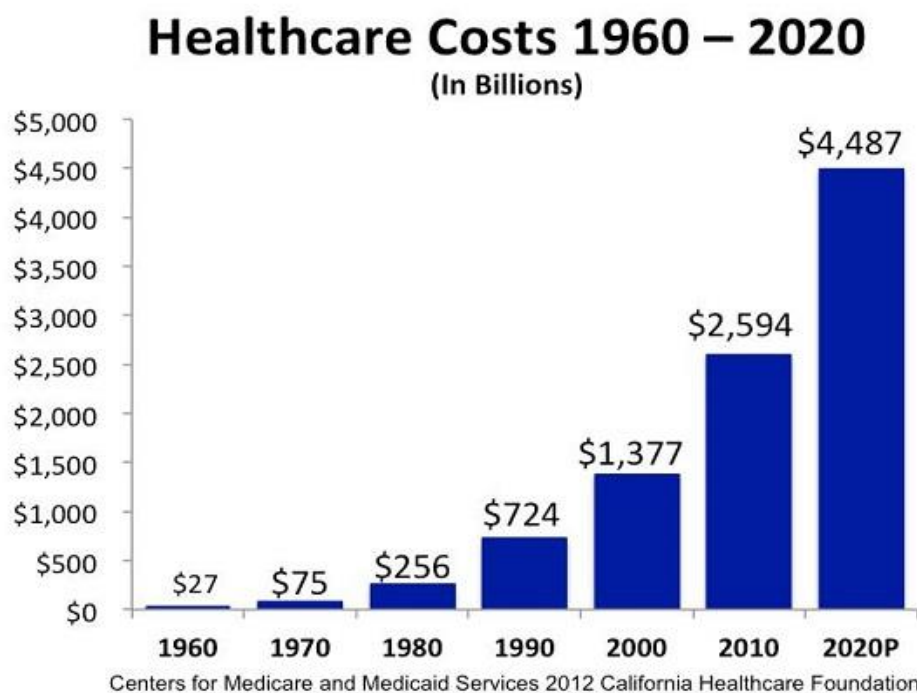


Εικόνα 3.12: Απεικόνιση πλήθους διαβητικών ατόμων ανά περιοχή στην Πολιτεία της Βαλτιμόρης

Chartio

Το Chartio προσφέρει τη δυνατότητα συνδυασμού των πηγών δεδομένων και της εκτέλεσης ερωτημάτων (queries) στο πρόγραμμα περιήγησης. Το σημαντικότερο πλεονέκτημα της εφαρμογής αυτής είναι η ταχύτητα με την οποία μπορεί να λειτουργήσει, καθώς σε ελάχιστο χρόνο μπορεί να εισάγει δεδομένα από διαφορετικές πηγές και να τα αξιοποιήσει χωρίς να απαιτούνται γνώσεις SQL ή άλλων πολύπλοκων γλώσσες προγραμματισμού. Χρησιμοποιείται συνήθως για την εξαγωγή απλούστερων διαγραμμάτων.

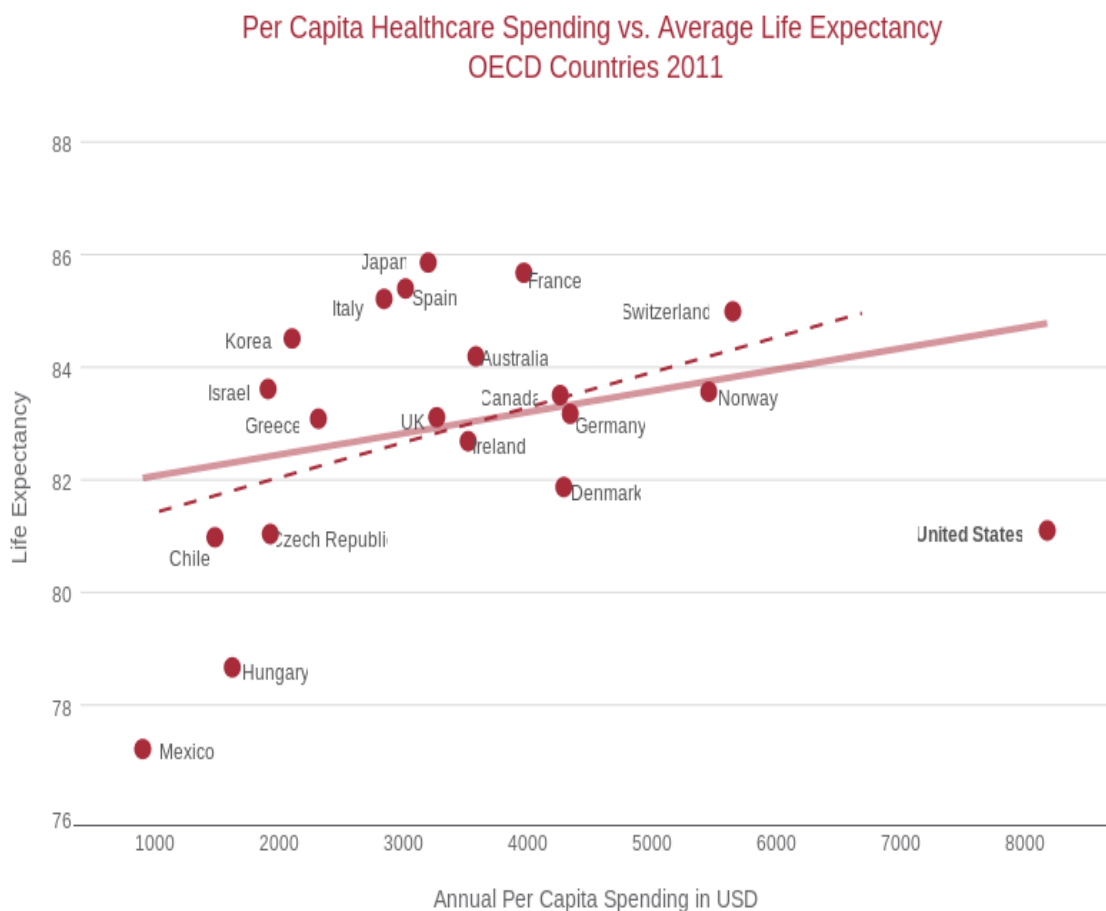
Στην ακόλουθη εικόνα, απεικονίζεται το συνολικό κόστος σε δισεκατομμύρια δολάρια των υπηρεσιών υγείας στις ΗΠΑ από το ίδρυμα California Healthcare Foundation, από το 1960 έως και το 2010. Παρουσιάζεται επίσης η αντίστοιχη πρόβλεψη, για το 2020.



Εικόνα 3.13: Συνολικό κόστος σε δισεκατομμύρια δολάρια των υπηρεσιών υγείας στις ΗΠΑ προς το χρόνο

Plot.ly

Το τελευταίο σχεδιαστικό εργαλείο που θα παρουσιαστεί είναι το Plot.ly, το οποίο είναι επίσης εύκολο στη χρήση, δίχως να απαιτεί τεχνικές γνώσεις, και μέσω αυτού γίνεται άμεσα να δημιουργηθούν δυσδιάστατα, αλλά και τρισδιάστατα γραφήματα, μία δυνατότητα που δεν είναι εξίσου εύκολο να υλοποιηθεί στις προηγούμενες εφαρμογές. Ακολουθεί τρισδιάστατη απεικόνιση του προσδόκιμου ζωής και του κατά κεφαλήν κόστους για υπηρεσίες και προϊόντα υγείας σε διάφορα κράτη, σε ένα ενιαίο γράφημα.



Εικόνα 3.14: Τρισδιάστατη απεικόνιση προσδόκιμου ζωής και κατά κεφαλήν κόστους για υπηρεσίες υγείας ανά κράτος

3.2.4 Γλώσσες Προγραμματισμού

Για την επεξεργασία και ανάλυση των Big Data, υπάρχει ανάγκη επιλογής των κατάλληλων γλωσσών προγραμματισμού. Οι κυριότερες γλώσσες που χρησιμοποιούνται στα Big Data Analytics είναι οι R, Python και Scala, ενώ σε κάποιες εφαρμογές χρησιμοποιούνται οι Java, RegEx και XPath.

Python

Η Python είναι μια υψηλού επιπέδου γλώσσα προγραμματισμού η οποία δημιουργήθηκε το 1990.²⁷ Ο κύριος στόχος της είναι η αναγνωσιμότητα του κώδικά της και η ευκολία χρήσης της. Το συντακτικό της επιτρέπει στους προγραμματιστές να εκφράσουν έννοιες σε λίγες γραμμές κώδικα. Διακρίνεται λόγω του ότι έχει πολλές βιβλιοθήκες που διευκολύνουν ιδιαίτερα αρκετές εργασίες και για την ταχύτητα εκμάθησής της.

R

Η R είναι γλώσσα προγραμματισμού και περιβάλλον που παρέχει στον χρήστη τη δυνατότητα να κάνει υπολογιστική στατιστική και γραφήματα.²⁸ Παρέχει τα απαραίτητα εργαλεία προκειμένου να υλοποιηθεί μια στατιστική ανάλυση, όπως:

- δημιουργία τυχαίων δειγμάτων
- διακριτές και συνεχείς μεταβλητές (Poisson, Gamma, Exponential)
- έλεγχοι υποθέσεων
- στατιστικά τεστ (Kolmogorov-Smirnoff)
- δημιουργία γραφημάτων (ιστόγραμμα, qq plot, pie chart, bar chart)

Scala

Η Scala είναι μια γλώσσα προγραμματισμού πολλαπλών παραδειγμάτων που σχεδιάστηκε για να ενσωματώσει χαρακτηριστικά του αντικειμενοστρεφούς και του συναρτησιακού προγραμματισμού.²⁹ Το όνομα Scala προέρχεται από την αγγλική φράση "scalable language", που δηλώνει ότι έχει σχεδιαστεί για να μπορεί να μεγαλώνει παράλληλα με τις ανάγκες των χρηστών της.

Java

Η Java είναι μια αντικειμενοστρεφής γλώσσα προγραμματισμού που σχεδιάστηκε από την εταιρεία πληροφορικής Sun Microsystems. Ένα από τα βασικά πλεονεκτήματά της έναντι των περισσότερων άλλων γλωσσών είναι η ανεξαρτησία του λειτουργικού συστήματος και της πλατφόρμας.³⁰

3.3 Εμπορικές εφαρμογές

Η εταιρία **IBM** παρέχει στα πλαίσια εμπορικής χρήσης την πλατφόρμα **Info Sphere**, η οποία παρέχει τις εξής υπηρεσίες και εφαρμογές:

- Το Apache Hadoop, το οποίο επεξεργάζεται και αναλύει όλων των τύπων τα δεδομένα που μπορεί να βρίσκονται σε συστάδες εξυπηρετητών (server clusters).
- Τρόπο αποθήκευσης δεδομένων (data warehouse) για βελτιωμένη αναζήτηση βάσει της λειτουργίας και των αναγκών της επιχείρησης.
- Ειδικά σχεδιασμένο λογισμικό για ανάλυση μεγάλου όγκου δεδομένων σε πραγματικό χρόνο.
- Υποστηρικτικές υπηρεσίες (εργαλεία διαχείρισης του συστήματος, περιβάλλον ανάπτυξης) και εργαλεία επιχειρηματικής ευφυΐας και ανάλυσης.¹⁷

Η εταιρία **Cisco**, έχει δημιουργήσει την πλατφόρμα **Cisco Common Big Data Platform**, βασισμένη στο Cisco's Unified Computing System. Η συγκεκριμένη πλατφόρμα προσφέρει δυνατότητες επεξεργασίας, δικτύωσης, αποθήκευσης και συνεργάζεται με μεγάλο πλήθος εμπορικών προϊόντων (π.χ. Datasta, Hortonworks, Cloudera, Oracle, Intel και Pivotal).¹²

Η **Amazon** προσφέρει επίσης λύσεις για Big Data μέσω του **Elastic MapReduce (EMR)**, που είναι επίσης βασισμένο στο Hadoop. Μέσω των **Amazon Web Services (AWS)**, είναι δυνατή η χρήση του **MapR M7** πάνω στην πλατφόρμα EMR. Επίσης η **Amazon** ανέπτυξε και δύο ακόμη προϊόντα, τη NoSQL βάση δεδομένων **DynamoDB** και ένα σύστημα διαχείρισης βάσεων δεδομένων, το **RedShift**. Το Amazon EMR συνεργάζεται με το Hadoop μέσω του AWS Cloud. η πλατφόρμα **MapR** επιτρέπει υψηλές ταχύτητες πρόσβασης στο Hadoop και σε NoSQL βάσεις δεδομένων, ενώ το M7 επιτρέπει την εύκολη κλιμάκωση. Το Amazon Redshift είναι μια υπηρεσία διαχείρισης βάσεων δεδομένων μεγέθους κάποιων εκατοντάδων gigabytes μέχρι και petabytes. Υποστηρίζει την ανάλυση των δεδομένων των επιχειρήσεων μέσω δεδομένων εργαλείων επιχειρηματικής ευφυΐας. Τέλος, το **Elastic Compute Cloud (Amazon EC2)** προσφέρει άμεση αλλαγή του απαιτούμενου μεγέθους της υπολογιστικής ισχύος στο νέφος.³¹

Το **Cloudera** είναι από τις κυρίαρχες λύσεις για το Apache Hadoop.³² Η εταιρία προσφέρει επίσης την εφαρμογή **Impala** για ταχύτερη επεξεργασία, μία μηχανή αναζήτησης η οποία σε πραγματικό χρόνο επιτρέπει την αναζήτηση δεδομένων αποθηκευμένων σε Hadoop Distributed File System (HDFS), καθώς και σε πίνακες βάσεων δεδομένων βασισμένων σε Apache HBase μέσα σε λίγα δευτερόλεπτα, μέσω μιας εύχρηστης διεπαφής για SQL ερωτήματα (queries). Η έκδοση Enterprise Cloudera παρέχει επίσης τη δυνατότητα διασύνδεσης μέσω ODBC (Open Database Connectivity) και JDBC (Java Database Connectivity).¹⁹

Η **Datameer** μέσω περιβάλλοντος βασισμένου σε φύλλα εργασίας, αναλύει τα δεδομένα, ενώ δίνει τη δυνατότητα οπτικής απεικόνισης των δεδομένων μέσω εικόνας, αποτελώντας μία ιδιαίτερα φιλική προς το χρήστη εφαρμογή. Βασίζεται επίσης στο Apache Hadoop.⁸

Η **DataStax** χρησιμοποιεί τη βάση δεδομένων Apache Cassandra ως NoSQL βάση δεδομένων και δίνει τη δυνατότητα επεξεργασίας και ανάλυσης δεδομένων, είτε αυτά είναι σε αυστηρά δομημένη είτε σε μη δομημένη μορφή, στο νέφος (cloud) . Η έκδοση **DataStax Enterprise** μάλιστα μέσω του εργαλείου **Apache Solr** βελτιώνει την ταχύτητα και την αποδοτικότητα της ανάλυσης των δεδομένων.³³

Τέλος, οι πλατφόρμες **Karmasphere** , **MapR M7**, **Platfora**, **Splunk** και **Hadapt** αποτελούν ολοκληρωμένα εργαλεία για αξιοποίηση και ανάλυση των Big Data.

Είναι ιδιαίτερα σημαντικό να αναφερθεί πως οι επιλογές ανοικτού κώδικα, αν και παρέχουν τη δυνατότητα ελεύθερης χρήσης, συνήθως απαιτούν εξειδίκευση, τεχνογνωσία και αρκετό χρόνο ώστε να καταστούν λειτουργικές για αποτελεσματική ανάλυση των διαθέσιμων δεδομένων. Για το λόγο αυτό, οι περισσότερες εταιρίες και οργανισμοί καταφεύγουν στις λύσεις που αναφέρθηκαν.

ΚΕΦΑΛΑΙΟ 4

ΣΥΜΒΟΛΗ ΣΤΗΝ ΥΓΕΙΑ

4.1 Πεδία δράσης και ευκαιρίες

Ο τομέας της υγείας απασχολεί σήμερα το 8% του συνολικού ευρωπαϊκού εργατικού δυναμικού και αποτελεί το 10% του Ακαθάριστου Εγχώριου Προϊόντος της Ευρωπαϊκής Ένωσης. Παράλληλα, οι δημόσιες δαπάνες για την υγειονομική περίθαλψη και τη μακροχρόνια φροντίδα αναμένεται να αυξηθούν κατά ένα τρίτο έως το 2060.³⁴

Αυτό οφείλεται κυρίως στην ταχεία γήρανση του πληθυσμού, στην αυξημένη παρουσία των χρόνιων ασθενειών και στις εξελίξεις της Ιατρικής τεχνολογίας, των οποίων η εφαρμογή είναι ιδιαίτερα δαπανηρή. Το σχετικά μεγάλο μερίδιο των δημοσίων δαπανών υγείας στο σύνολο των δημοσίων δαπανών, σε συνδυασμό με την ανάγκη εξυγίανσης των δημοσιονομικών υπηρεσιών, καθώς και του εξορθολογισμού του προϋπολογισμού σε ολόκληρη την Ευρωπαϊκή Ένωση, επιβεβαιώνουν την ανάγκη βιωσιμότητας των σύγχρονων συστημάτων υγείας. Τα στοιχεία του Οργανισμού Οικονομικής Συνεργασίας και Ανάπτυξης (Ο.Ο.Σ.Α.) δείχνουν ότι με τη βελτίωση της παραγωγικότητας του συστήματος υγειονομικής περίθαλψης, η εξοικονόμηση δημοσίων δαπανών θα είναι τόσο μεγάλη, ικανή να προσεγγίσει το 2% του Ακαθάριστου Εγχώριου Προϊόντος κατά μέσο όρο στα ευρωπαϊκά κράτη, το οποίο θα ισοδυναμούσε με 330 δισεκατομμύρια ευρώ συνολικά στην Ευρώπη.³⁵

Οι τεχνολογίες των Big Data έχουν ήδη διαμορφώσει εξελίξεις σε τομείς που σχετίζονται με την υγειονομική περίθαλψη: Ιατρική διάγνωση από δεδομένα απεικόνισης στην Ιατρική, ποσοτικοποίηση των δεδομένων του τρόπου ζωής για τη βελτίωση της φυσικής κατάστασης και πολλά ακόμα παραδείγματα επιβεβαιώνουν αυτό τον ισχυρισμό. Παρ' όλα αυτά, η υγειονομική περίθαλψη παρουσιάζει σχετική καθυστέρηση στην αξιοποίηση των εξελίξεων των Big Data, η οποία είναι μια παράδοση κατάσταση, δεδομένου ότι ήδη από το 2012, το Ινστιτούτο Poneman, εκτιμούσε πως το 30% όλων των ηλεκτρονικών δεδομένων παγκοσμίως αποτελούνταν από δεδομένα σχετικά με την υγειονομική περίθαλψη.³⁶ Είναι προφανές ότι στο πλαίσιο του υφιστάμενου όγκου των μεγάλων δεδομένων υπάρχει κρυμμένη γνώση που θα μπορούσε να αλλάξει τη ζωή όχι μόνο ενός ασθενούς, αλλά σε πολύ μεγάλο βαθμό, να αλλάξει ο ίδιος ο κόσμος. Η εξαγωγή αυτής της γνώσης είναι ο πιο γρήγορος, λιγότερο δαπανηρός και πιο αποτελεσματικός τρόπος για τη βελτίωση της ανθρώπινης υγείας σε μεγάλη κλίμακα.³⁷

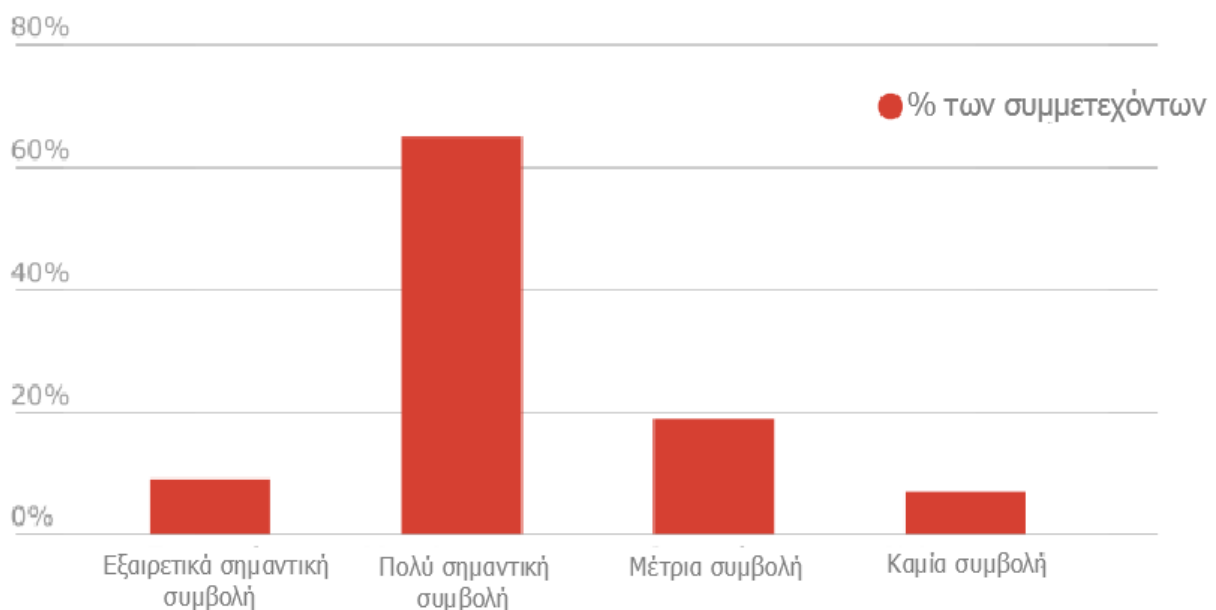
Η πρόοδος των εξελίξεων που συμβαίνουν συνεχώς στον τομέα της αξιοποίησης των Big Data, θα δημιουργήσει αναπόφευκτα νέες ευκαιρίες και θα επιτρέψει καινοτομίες που σχετίζονται με τη

διαχείριση πάσης φύσεως δεδομένων υγειονομικής περίθαλψης, υπό διαφορετικές οπτικές γωνίες. Συγκεκριμένα, η Ιατρική επιστήμη θα είναι σε θέση να:

- Περιγράψει καλύτερα τις διαδικασίες που πραγματοποιούνται στον οργανισμό κατά τη διάρκεια μιας ασθένειας, εφόσον υπάρχει άμεση πρόσβαση στο ανθρώπινο γονιδίωμα.
- Διαγνώσει αποτελεσματικότερα μια ασθένεια και τα αίτιά της.
- Προβλέψει τον τρόπο που θα εξελιχθεί μια ασθένεια.
- Προτείνει τρόπους για την αντιμετώπιση και τον περιορισμό μιας ασθένειας.

Από τα παραπάνω, είναι πέραν αμφιβολίας ότι ο πιθανός αντίκτυπος της τεχνολογίας των Big Data στην υγεία είναι τεράστιος. Υπάρχει ήδη ισχυρή τάση για καινοτομίες από οργανισμούς και φορείς. Μάλιστα σε συνέδριο που πραγματοποιήθηκε το 2015 στο Σαν Χοσέ, περισσότερο από το 60% των παρευρισκομένων επαγγελματιών υγείας δήλωσαν πως αναμένουν πολύ σημαντική μελλοντική συμβολή των Big Data στον κλάδο της υγείας, ενώ το 10% θεωρεί πως η συμβολή θα είναι εξαιρετικής σημασίας.³⁸

Εκτίμηση μελλοντικής συμβολής των Big Data στον τομέα της υγείας από επαγγελματίες υγείας

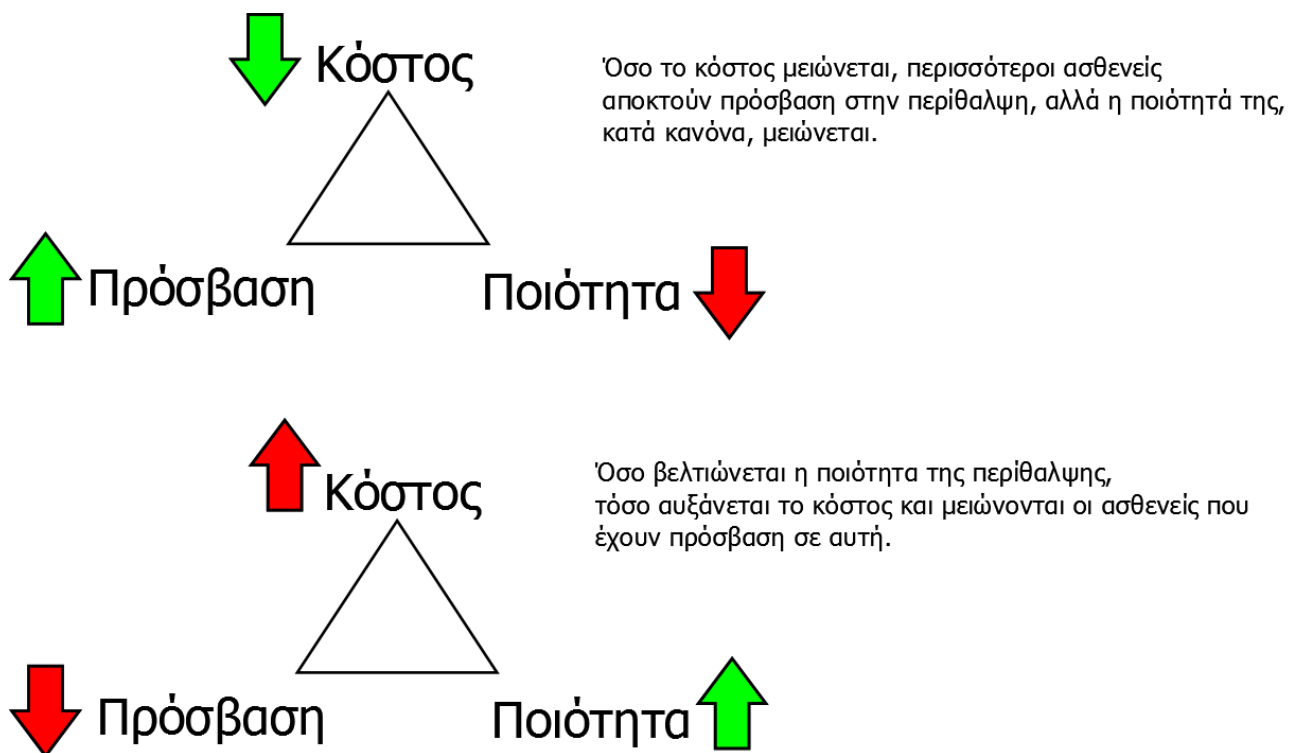


Εικόνα 4.1: Εκτίμηση μελλοντικής συμβολής των Big Data στον τομέα της υγείας από επαγγελματίες υγείας

4.2 Οικονομικές προοπτικές

Η ταχεία γήρανση του πληθυσμού συμβάλλει στις διαρκώς αυξανόμενες απαιτήσεις καθώς οι χρόνιες ασθένειες είναι πιο διαδεδομένες στον ηλικιωμένο πληθυσμό. Μόνο στην Ευρώπη, ο αριθμός των ανθρώπων ηλικίας 85 ετών και άνω προβλέπεται να αυξηθεί από 14 έως 19 εκατομμύρια παγκοσμίως έως το 2020 και κατά 40 εκατομμύρια μέχρι το 2050.³⁹ Η επίδραση αυτών των συνεχώς αυξανόμενων απαιτήσεων φαίνεται καθαρά από μια μελέτη που διεξήχθη από την εταιρία Accenture το 2014 η οποία διαπίστωσε ότι ένα τρίτο των ευρωπαϊκών νοσοκομείων είχε αναφέρει λειτουργικές απώλειες λόγω του φαινομένου.⁴⁰ Αυτό επιβεβαιώνει το γεγονός ότι οι χώρες της Ευρώπης αντιμετωπίζουν συνεχώς αυξανόμενες δυσκολίες, στην προσπάθειά τους να παρέχουν υψηλού επιπέδου υπηρεσίες υγείας σε χαμηλό κόστος.⁴¹

Ο όρος **Σιδηρούν Τρίγωνο της Υγείας (Iron Triangle of Healthcare⁴²)** χρησιμοποιείται για να περιγράψει αυτή την πρόκληση. Οι τρεις συνιστώσες του τριγώνου είναι η **ποιότητα (quality)**, η **πρόσβαση (access)** και το **κόστος (cost)**. Η αποτελεσματικότητα, η αξία και η έκβαση της φροντίδας αντικατοπτρίζει την ποιότητα του συστήματος υγειονομικής περίθαλψης. Η πρόσβαση είναι ενδεικτική για το πλήθος των πολιτών που μπορούν να λάβουν περίθαλψη όταν τη χρειάζονται. Το κόστος αντιπροσωπεύει την τιμή της περίθαλψης και την οικονομική προσιτότητα των ασθενών και των φορέων διαχείρισης μονάδων υγείας. Το πρόβλημα είναι ότι οι τρεις δομικοί λίθοι του τριγώνου βρίσκονται σε ανταγωνισμό μεταξύ τους στον τομέα της υγείας. Έτσι, ενώ μπορεί να είναι δυνατόν να υπάρξει βελτίωση ενός ή δύο συστατικών, στις περισσότερες περιπτώσεις, αυτό έρχεται σε βάρος του τρίτου, όπως φαίνεται στο ακόλουθο σχήμα.



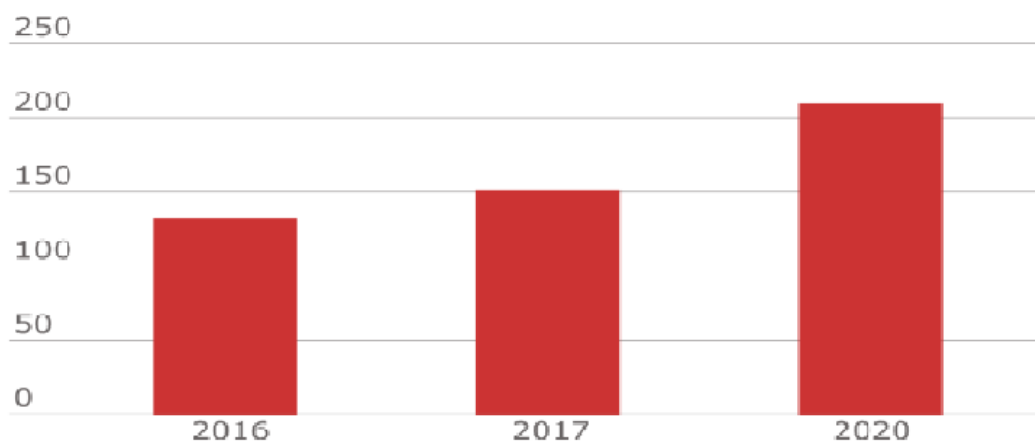
Εικόνα 4.2: Το Σιδηρούν Τρίγωνο της Υγείας (Iron Triangle of Healthcare)

Ωστόσο, ενώ οι παρούσες προσεγγίσεις βελτιστοποίησης της υγείας μπορούν να βοηθήσουν στην προσθήκη μικρών αλλαγών στην ισορροπία, για το Σιδηρούν Τρίγωνο της Υγείας, μόνο μια ριζική επανάσταση έχει τη δυνατότητα να προκαλέσει αλλαγές, τέτοιες ώστε και τα τρία συστατικά μέρη να βελτιωθούν ταυτόχρονα. Δεδομένου ότι η υγειονομική περίθαλψη αποτελεί μία από τις πιο ισχυρές βιομηχανίες παραγωγής Big Data, λόγω του υψηλού όγκου, της μεγάλης ποικιλίας, της υψηλής ακρίβειας και της αξίας των πηγών δεδομένων, τα Big Data Analytics έχουν τη δυνατότητα να ανατρέψουν τον κανόνα ισορροπίας στο Σιδηρούν Τρίγωνο. Αν και τα περισσότερα από τα δεδομένα παραδοσιακά ήταν σε μη αξιοποιήσιμη μορφή για τα υπολογιστικά συστήματα, υπάρχει έντονη προσπάθεια ψηφιοποίησής τους, η οποία τα καθιστά υπερπολύτιμα.

Ο τομέας των Big Data αποτελεί ένα χώρο στον οποίο οι επενδύσεις συνεχώς αυξάνονται. Σύμφωνα με έρευνες του φορέα International Data Corporation, το χρηματικό ποσό που αναμένεται να δαπανηθεί παγκοσμίως για την ανάλυση Big Data, είναι της τάξεως των 150,8 δισεκατομμυρίων δολλαρίων, παρουσιάζοντας αύξηση 12% σε σχέση με το 2016.⁴³ Επίσης, βάσει της ίδιας πηγής, προβλέπεται διατήρηση του ετήσιου ρυθμού αύξησης των επενδύσεων στο 12%, έως και το 2020, οπότε το συνολικό ύψος των επενδύσεων θα αγγίξει τα 210 δισεκατομμύρια δολάρια.⁴⁴

Πορεία του ύψους των επενδύσεων σε τεχνολογίες Big Data Analytics παγκοσμίως

(σε δισεκατομμύρια δολάρια)



Εικόνα 4.3: Επενδύσεις σε δισεκατομμύρια δολάρια για Big Data Analytics παγκοσμίως ως προς το χρόνο

Ταυτόχρονα, η παγκόσμια αγορά για ανάλυση δεδομένων υγείας αναμένεται να φτάσει τα 18,7 δισεκατομμύρια δολάρια το 2020 από τα 5,8 δισεκατομμύρια δολάρια που ήταν το 2015.⁴⁵ Δικαιολογημένα αναμένεται μεγάλο μέρος των επενδύσεων να αξιοποιηθεί στον τομέα της υγείας. Αυτό έρχεται ως φυσικό επακόλουθο του δυνητικού κέρδους που μπορεί να προκύψει τόσο σε ποιοτικό επίπεδο για την υγεία, όσο και σε επίπεδο εξοικονόμησης πόρων. Στην ακόλουθη εικόνα παρουσιάζονται τα χρηματικά ποσά που αναμένεται να εξοικονομηθούν σε ετήσια βάση, μόνο στις ΗΠΑ, μέσω της αξιοποίησης της τεχνολογίας των Big Data σε διάφορα πεδία, σε μία έρευνα της εταιρίας McKinsey.¹⁴

Προοπτικές εξοικονόμησης πόρων στις ΗΠΑ σε δισεκατομμύρια δολάρια ετησίως

Πρόληψη ασθενειών βάσει δεδομένων

\$100

Υιοθέτηση αποδεδειγμένων μεθόδων θεραπείας από συγκεντρωτικά δεδομένα οργανισμών υγείας

\$110

Μειωμένη χρήση Μονάδων Εντατικής Θεραπείας

\$70

Επίτευξη διαφάνειας στις οικονομικές δόσοληψίες

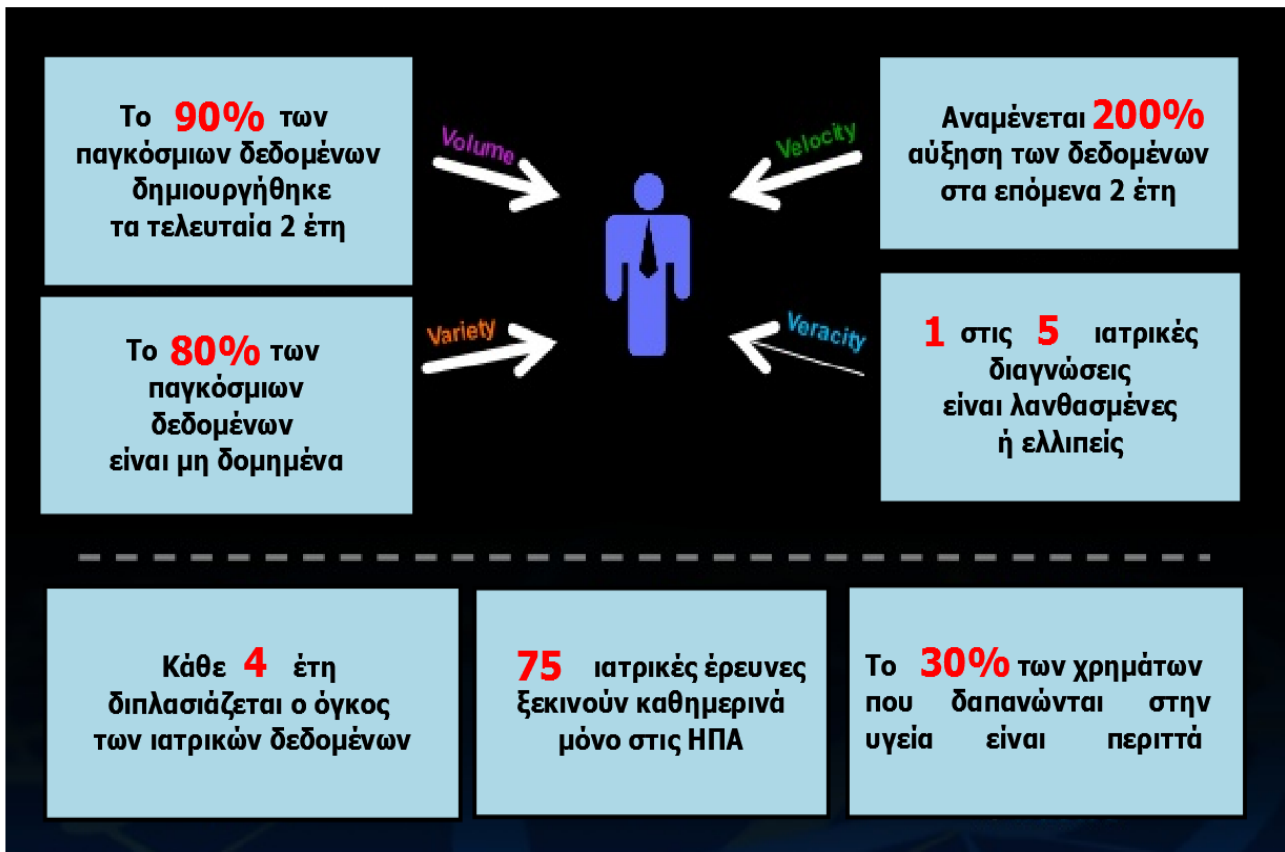
\$100

Στήριξη της έρευνας και επιτάχυνση των καινοτομιών

\$70

Εικόνα 4.4: Δυνατή εξοικονόμηση πόρων στις ΗΠΑ σε δισεκατομμύρια δολάρια από χρήση Big Data Analytics στην υγεία

Στην ακόλουθη εικόνα παρουσιάζεται με αριθμούς η ανάγκη αξιοποίησης των Big Data στον τομέα της υγείας, με στοιχεία της εταιρίας IBM:⁴⁶



Εικόνα 4.5: Η ανάγκη αξιοποίησης των Big Data στον τομέα της υγείας με αριθμούς

Προς υποστήριξη της χρησιμότητας των Big Data στην υγεία, καταγράφεται η προσπάθεια τήρησης προσωπικών ηλεκτρονικών ιατρικών αρχείων, ως μέσο αξιοποίησης των δεδομένων.

4.3 Ηλεκτρονικός Φάκελος Υγείας (EHR)

Ο **Ηλεκτρονικός Φάκελος Υγείας (Electronic Health Record ή EHR)**, ο οποίος συχνά αναφέρεται και ως **Ηλεκτρονικός Ιατρικός Φάκελος (Electronic Medical Record ή EMR)**, ορίζεται ως: «Ένα αποθετήριο πληροφοριών σχετικών με την υγεία ενός ατόμου υπό φροντίδα, σε επεξεργάσιμη μορφή από υπολογιστή που αποθηκεύεται και μεταφέρεται με ασφάλεια. Είναι εύκολα προσβάσιμο από τους εξουσιοδοτημένους χρήστες. Έχει ένα τυποποιημένο μοντέλο πληροφοριών, το οποίο είναι ανεξάρτητο από τα συστήματα που το χρησιμοποιούν. Πρωταρχικός σκοπός του είναι η υποστήριξη μιας συνεχούς, ποιοτικής, ολοκληρωμένης υγειονομικής περίθαλψης και η παροχή πληροφοριών».⁴⁷

Οι βασικοί στόχοι που εξυπηρετεί ο Ηλεκτρονικός Φάκελος Υγείας, είναι οι ακόλουθοι:

- Πρόσβαση σε δεδομένα υγείας και πληροφορίες για τους ασθενείς
- Αλληλεπίδραση μεταξύ των μονάδων υγείας
- Υποστήριξη των ασθενών
- Εποπτεία των διαδικασιών του τομέα υγειονομικής περίθαλψης
- Δυνατότητα εύκολης και άμεσης επεξεργασίας δεδομένων
- Βελτίωση του τρόπου λήψης αποφάσεων σχετικών με τη θεραπεία
- Αποδοτικότερη διαχείριση των υπηρεσιών υγείας

Μέσω των αρχείων αυτών πρόσβαση στην πληροφορία παρέχεται σε:

- Ασθενείς
- Διοικητικό προσωπικό
- Εργαστηριακό προσωπικό
- Θεράποντες ιατρούς
- Νοσηλευτικό προσωπικό
- Τμήματα πληροφορικής οργανισμών
- Προσωπικό φαρμακείων

Σύμφωνα με τον **HIMSS** (Healthcare Information and Management Systems Society), η διαλειτουργικότητα στην υγειονομική περίθαλψη ορίζεται ως η ικανότητα διαφορετικών εφαρμογών λογισμικού και συστημάτων πληροφορικής να ανταλλάσσουν δεδομένα, να επικοινωνούν και να χρησιμοποιούν τις πληροφορίες που έχουν ανταλλάξει. Το σχήμα ανταλλαγής δεδομένων και τα

πρότυπα και προδιαγραφές που χρησιμοποιούνται, πρέπει να επιτρέπουν την ανταλλαγή δεδομένων ανάμεσα σε ιατρούς, εργαστήρια, νοσοκομεία, φαρμακεία, και ασθενείς, ανεξάρτητα από την περίπτωση ή τον προμηθευτή.⁴⁸

Η δυνατότητα αξιοποίησης των Big Data, στηρίζεται σε τεράστιο βαθμό στη δυνατότητα πρόσβασης στα δεδομένα και στη διαλειτουργικότητα των πληροφοριακών συστημάτων υγείας. Συναντώνται τρία επίπεδα διαλειτουργικότητας:

- **Θεμελιώδης διαλειτουργικότητα:** Αφορά την ικανότητα μεταφοράς των δεδομένων μεταξύ των συστημάτων.
- **Διαρθρωτική λειτουργικότητα:** Ορίζει τα πρότυπα δομής των δεδομένων και του τρόπου μεταφοράς τους.
- **Σημασιολογική διαλειτουργικότητα:** Αφορά την ικανότητα αξιοποίησης των δεδομένων, αφού έχουν μεταφερθεί από ένα σύστημα σε ένα άλλο.

Στην ακόλουθη εικόνα απεικονίζεται μία τυπική περίπτωση Ηλεκτρονικού Φακέλου Υγείας ασθενούς, με τα συγκεντρωτικά στοιχεία που τον αφορούν:

The screenshot displays a comprehensive patient record for James Noble. Key sections include:

- Appointments:** A list of upcoming and past appointments with times and names of providers.
- Clinical Summary:** Patient name, address (4516 West Huron Street, Chicago, IL 60607), primary insurance (Aetna U.S. Healthcare - Master), and pharmacy directions.
- Demographics:** Patient photo, email (jnobles@gmail.com), DOB (04.05.1962), age (50y), and patient ID (88501).
- User Defined Fields:** Personal details such as injury (Shoulder pain), employer (WBI), insurance carrier (GEICO), and position (Truck Driver).
- Diagnosis:** Current and past diagnoses including ACL Tear (06.20.2012) and Lumbago (11.17.2011).
- Vitals & Smoking Status:** Vital signs and smoking history (Never smoker).
- Procedures:** A list of performed procedures, including Knee Arthroscopy/Surg on 06.20.2012.
- Rx History:** A table of medications:

Status	Date	Drug	Strength	Instructions
Active	06.18.2012	Keflex	250 MG	1 tab po BID
Active	11.04.2010	CeleBREX	100 MG	1 tablet by mouth in the AM
Active	04.13.2006	OxyCODONE HCl	20 MG	1 tab po q4h prn pain
- Non-Drug Allergies:**

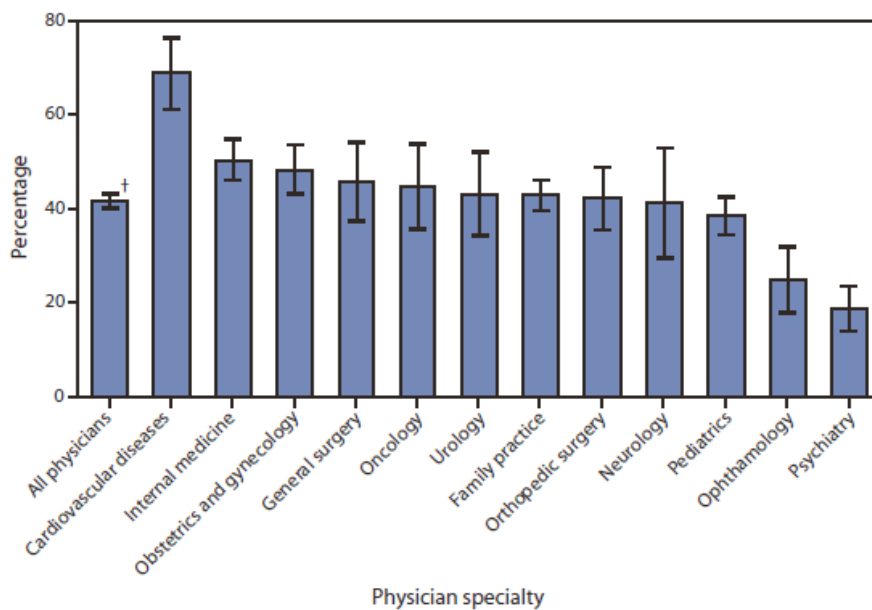
Description	Reaction	Notes
Latex	Severe rash	
Shellfish	Hives	
- Family History:**

Relationship	Deceased	Notes
Sister	No	JFA
Maternal Grandmother	Yes	Osteoarthritis
- Surgeries:**

Description	Date	Surgeon	Notes
Arthroscopy	12.18.2011	Armstrong	N/A
- Appointments:** A table of scheduled appointments with columns for Date, Time, Doctor, Reason, Type, Location, and Notes.

Εικόνα 4.6: Τυπική περίπτωση Ηλεκτρονικού Φακέλου Υγείας ασθενούς

Η χρήση του Ηλεκτρονικού Φακέλου Υγείας, είναι πλέον ιδιαίτερα διαδεδομένη. Παρ' όλα αυτά, υπάρχει σχετική διακύμανση μεταξύ του ποσοστού των επαγγελματιών ιατρών που την έχουν υιοθετήσει, ανάλογα με την ειδικότητα, όπως επιβεβαιώνεται και από την έρευνα που έγινε το 2012 στις ΗΠΑ:⁴⁹



Εικόνα 4.7: Χρήση ΗΦΥ ανά ειδικότητα ιατρών

Ως συμπέρασμα, προκύπτει πως οι ίδιοι λόγοι που οδήγησαν στη δημιουργία του Ηλεκτρονικού Φακέλου Υγείας, προκάλεσαν και τις εξελίξεις στην τεχνολογία των Big Data. Τα πολλαπλά οφέλη που μπορεί να προσφέρει η αξιοποίηση των διαθέσιμων δεδομένων στην υγεία, τόσο σε επίπεδο ποιότητας, όσο και στον περιορισμό των δαπανών, έχουν διαμορφώσει τις εξελίξεις. Ουσιαστικά ο Ηλεκτρονικός Φάκελος Υγείας αποτελεί την πρώτη συλλογική προσπάθεια συγκεντρωτικής συλλογής των δεδομένων και συνεισφέρει σημαντικά στην πρόσβαση και επεξεργασία των Big Data. Η ειδοποιός διαφορά έγκειται στο ότι τα Big Data Analytics, αν και αξιοποιούν τα διαθέσιμα δεδομένα, δεν απαιτούν τη συγκέντρωσή τους υπό ενιαία δομή σε μία βάση, παρέχοντας ουσιαστικά ένα υπερσύνολο δυνατοτήτων από εκείνων του Ηλεκτρονικού Φακέλου Υγείας.

4.4 Οι ευκαιρίες με το μεγαλύτερο αντίκτυπο

Αυτή η ενότητα περιγράφει συγκεκριμένες περιοχές στον τομέα της υγείας που θα ωφεληθούν περισσότερο από την εφαρμογή των τεχνολογιών αξιοποίησης των Big Data.

4.4.1 Υγιεινή Ζωή και Πρόληψη: Αλλαγή νοοτροπίας

Η ευρεία χρήση των Big Data θα μπορούσε να συμβάλλει στην παροχή αποτελεσματικών εργαλείων για την αλλαγή της νοοτροπίας και συμπεριφοράς των ανθρώπων με στόχο έναν πιο υγιεινό τρόπο διαβίωσης. Ειδικά η χρήση κινητών τηλεφώνων στην υγεία (Mobile Health ή mHealth) έχει τη δυνατότητα να προσαρμόσει τις προτάσεις σε κάθε άτομο, αξιοποιώντας τα δεδομένα του τρόπου ζωής (διατροφή, σωματική άσκηση, ύπνος) και να εκτιμήσει τα δεδομένα επιρροής σε επιμέρους πληθυσμιακές ομάδες. Εκτός από την παροχή πληροφοριών για τους ανθρώπους, οι τεχνολογίες mHealth εκμεταλλεύονται συναφείς πληροφορίες που είναι το κλειδί της καινοτομίας. Με τον τρόπο αυτό, παρέχεται μια πλήρως ολοκληρωμένη εικόνα του τι επηρεάζει την πρόοδο και προκαλεί τις οπισθοδρομήσεις στη θεραπεία.

4.4.2 Κατανόηση των παραγόντων ενεργοποίησης των χρόνιων ασθενειών

Η καλύτερη κατανόηση της σχέσης μεταξύ κοινωνικών και ατομικών συμπεριφορών, όπως η διατροφή, οι γενετικοί και περιβαλλοντικοί παράγοντες είναι δυνατόν να δώσουν πολλές πληροφορίες σχετικά με την ανάπτυξη ψυχικών και σωματικών ασθενειών. Οι πολύπλοκες αλληλεπιδράσεις μεταξύ των διαφόρων συστημάτων που καθορίζουν την εξέλιξη μιας νόσου δεν είναι ακόμη πλήρως κατανοητές. Αναμένεται ότι η ολοκληρωμένη εικόνα της υγείας με βάση διάφορους δείκτες μπορεί να βοηθήσει και να βελτιώσει την έγκαιρη ανίχνευση των ασθενειών και μακροπρόθεσμα να συμβάλλει στη διαχείριση των αρνητικών παραγόντων για την υγεία μειώνοντας τα κόστη.

4.4.3 Περίθαλψη προσαρμοσμένη στις ανάγκες κάθε πληθυσμιακής ομάδας

Οι πολιτικές δημόσιας υγείας βασίζονται στην εκτενή ανάλυση των υγειονομικών δεδομένων που αφορούν πληθυσμούς, κυρίως με κριτήρια την κοινωνικο-οικονομική κατάσταση και τις περιοχές, ώστε να επιτυγχάνεται αποτελεσματική και συγκεκριμένη στόχευση για τη βελτίωση της

υγειονομικής περίθαλψης. Η ανάλυση των Big Data μπορεί να κατευθύνει τις πολιτικές με τέτοιο τρόπο ώστε να παρέχεται η κατάλληλη φροντίδα για κάθε πληθυσμιακή ομάδα. Η επιτυχία των εκάστοτε πολιτικών είναι συνάρτηση της ποιότητας των δεδομένων στα οποία στηρίζεται η έρευνα και στην προοπτική επιρροής των ενεργειών που προορίζονται για την εξυπηρέτησή τους.

Σε πολλές περιπτώσεις, δεν υπάρχουν κοινώς αποδεκτοί μέθοδοι για τον έλεγχο της εγκυρότητας των πολιτικών που εφαρμόζονται, γεγονός που οφείλεται σε παράγοντες όπως:

- Η προστασία των δεδομένων δυσχεραίνει τη σύγκριση πληροφορίας διαφορετικών οργανισμών υγείας, νοσοκομείων, ασφαλιστικών εταιριών κ.ό.κ.
- Το σημαντικότερο τμήμα των ιατρικών αρχείων είναι μη δομημένα δεδομένα
- Υπάρχουν δομικοί περιορισμοί ποιότητας και ενσωμάτωσης των δεδομένων
- Τα δημοφιλέστερα συστήματα διαχείρισης δεδομένων δεν επιτρέπουν τη δυναμική διαχείριση των Big Data.

4.4.4 Λοιμώδη νοσήματα

Η τεχνολογία των τελευταίων ετών επέτρεψε όχι μόνο τη λήψη δεδομένων από το περιβάλλον της υγειονομικής περίθαλψης (νοσοκομεία, κέντρα υγείας, εργαστήρια, κλπ), αλλά και πληροφορίες άμεσα προερχόμενες από τους ασθενείς (αισθητήρες, παρακολούθηση, συσκευές Internet of Things, κοινωνικά δίκτυα, κ.ό.κ.). Οι υπηρεσίες υγείας θα επωφεληθούν άμεσα μέσω της απόκτησης και της ανάλυσης των πληροφοριών που προκύπτουν από κάθε είδους προέλευση. Συνδυάζοντας τις πληροφορίες από ανεπίσημες πηγές, όπως αναζητήσεις στο διαδίκτυο και ιατρικά διαγνωστικά δεδομένα, συμπεριλαμβανομένων, για παράδειγμα, των διαφοροποιήσεων που παρατηρούνται σε αλληλουχίες γονιδιώματος, γίνεται δυνατή η παροχή έγκαιρης ανίχνευσης της αιτίας κάποιας νόσου, καθώς και λεπτομερείς πληροφορίες για την κατανόηση των μηχανισμών που την προκαλούν.⁵⁰

Χαρακτηριστικό παράδειγμα αποτελεί το μοντέλο ARGO, που χρησιμοποιεί διάφορες πηγές δεδομένων, συμπεριλαμβανομένων των δεδομένων αναζήτησης της Google, για να δημιουργήσει ένα μοντέλο πρόβλεψης για τη γρίπη.⁵¹ Διάφορα συστήματα έχουν δημιουργηθεί για την παρακολούθηση

των επιπέδων των ασθενειών, της δραστηριότητάς τους, ακόμα και τη δυναμική εξάπλωσή τους με τη χρήση πληροφοριών που παρέχονται από κοινωνικά δίκτυα. Χαρακτηριστικό παράδειγμα αποτελεί ο έλεγχος της εξάπλωσης της επιδημίας H1N1 μέσω του Twitter.⁵² Αναλύοντας τα στοιχεία αυτά σε συνδυασμό με πολλές μεταβλητές, όπως τα ταξίδια, το εμπόριο, οι κλιματικές αλλαγές κ.ό.κ., θα ήταν δυνατή η ανάπτυξη μοντέλων πρόβλεψης για παρεμβάσεις βασισμένες στον πληθυσμό, καθώς και η βελτιωμένη εξατομικευμένη θεραπεία του ασθενούς. Μέσω της χρήσης των Big Data θα ήταν επίσης εφικτή η ανίχνευση των οίωνων που προμηνύουν νέες λοιμώξεις και επιδημίες και η έγκαιρη επιβολή προληπτικών μέτρων.

4.4.5 Βελτίωση Ακρίβειας Ιατρικής Επιστήμης

Η συστηματική συλλογή και ανάλυση των γενετικών δεδομένων, σε συνδυασμό με πληροφορίες για τις ασθένειες, τις θεραπείες και τα αποτελέσματά τους έχει τη δυνατότητα να βελτιώσει αποφασιστικά την επιλογή των κατάλληλων θεραπειών, αποφεύγοντας τις επιπλοκές που παρουσιάζονται από λανθασμένους χειρισμούς και τη χρήση αναποτελεσματικών θεραπειών. Η διαθεσιμότητα των δεδομένων είναι δυνατόν να διαπιστώσει πιο αποτελεσματικά, εκτός από τα αίτια, και το στάδιο στο οποίο βρίσκεται μία ασθένεια, καθώς και την ακριβή κατάσταση του ασθενούς. Μία σημαντική νέα τεχνολογία που υπόσχεται βελτιώσεις στην ακρίβεια της Ιατρικής επιστήμης είναι η Υψηλής Απόδοσης Ανάλυση Γονιδιώματος (**High Performance Genome Analysis**). Η μεγάλη ποσότητα των γονιδιωματικών δεδομένων που είναι διαθέσιμα, δίνει τη δυνατότητα παραγωγής νέων αναλυτικών αλγορίθμων για ιατρική και κλινική χρήση. Θα γίνει έτσι εφικτή η σύγκριση ολόκληρων γονιδιωμάτων ασθενών με μεγάλα σύνολα πληθυσμού άλλων ατόμων. Η δημιουργία τεράστιου όγκου γονιδιωματικών βάσεων δεδομένων για σπάνιες ασθένειες είναι ένα τέτοιο παράδειγμα. Οι εφαρμογές αυτές είναι πολύπλοκες και απαιτητικές, επειδή χρειάζονται μη-συγκεντρωτικά δεδομένα (δεδομένα από πολλούς φορείς και πηγές) και μεγάλη υπολογιστική ισχύ.

4.4.6 Συλλογή αποτελεσμάτων περίθαλψης ασθενών και αντίστοιχων οικονομικών δαπανών

Βάσει της κατευθυντήριας αρχής για βιώσιμη υγειονομική περίθαλψη, τα αποτελέσματα από τη θεραπεία και τη φροντίδα των ασθενών οφείλουν να εξετάζονται παράλληλα με τις οικονομικές απαιτήσεις της μεθόδου που εφαρμόστηκε, ώστε να αξιολογηθούν επαρκώς.⁵³ Στον τομέα της υγείας λοιπόν, ο συνδυασμός κόστους - απόδοσης, είναι υψίστης σημασίας και καθορίζει τα κίνητρα των

φορέων παροχής υγειονομικής περίθαλψης για τη βελτίωση της ποιότητας και της αποτελεσματικότητας των υπηρεσιών. Με αυτήν την αρχή, τα δεδομένα πρέπει να συλλέγονται, να αναλύονται και να συγκεντρώνονται όντας ξεκάθαρη η σχέση μεταξύ της αποτελεσματικότητας και του κόστους. Συγκεκριμένα, τα αποτελέσματα της περίθαλψης που σχετίζονται με την υγεία του ασθενούς πρέπει να συλλέγονται και να επαληθεύονται πριν, κατά τη διάρκεια, και στη λήξη των θεραπειών, κάτι το οποίο αν και φαντάζει προφανές, δεν αποτελεί τη συνήθη πρακτική. Παράλληλα, είναι μια πρόκληση για την αναδιοργάνωση των συστημάτων διοίκησης μονάδων υγείας να επιτρέπουν τη σύνδεση όλων των εμπλεκόμενων δαπανών, προκειμένου να έχουν μια ακριβή εκτίμηση του πλήρους κόστους. Εφόσον οι διαδικασίες μπορέσουν να συνδεθούν, οι αποφάσεις για συγκεκριμένες θεραπείες μπορούν να βασίζονται όχι μόνο σε εμπειρικά στοιχεία, αλλά σε μια τεράστια βάση δεδομένων των ασθενών με παρόμοιες νόσους και των αντιστοίχων συνολικών δαπανών των θεραπειών που εφαρμόστηκαν. Είναι βέβαια απαραίτητο, από τη στιγμή που γίνει μία τέτοια προσπάθεια, να είναι οργανωμένη και οι εκτιμήσεις κόστους να γίνονται με ενιαία κριτήρια και πρότυπα ούτως ώστε να έχουν πραγματική αξία.

4.4.7 Βελτίωση της ροής των παραγωγικών μονάδων

Γενικά, στις περισσότερες μεγάλες βιομηχανίες και επιχειρήσεις, οι διαδικασίες που πραγματοποιούνται είναι προβλέψιμες σε μεγάλο βαθμό. Ωστόσο, οι διαδικασίες σε ένα νοσοκομείο παρουσιάζουν μεγάλη μεταβλητότητα και αλλάζουν διαρκώς, καθώς εξαρτώνται από πολλούς παράγοντες, όπως οι ανάγκες των εκάστοτε ασθενών, η εμπειρία και διαθεσιμότητα ιατρικού και νοσηλευτικού προσωπικού, ο τρόπος διοίκησης, η συνεργασία μεταξύ διαφορετικών τμημάτων κ.ό.κ.

Η φροντίδα της ανθρώπινης υγείας δεν αφήνει περιθώρια για πειραματισμούς, διαστήματα προσαρμογής και λανθασμένους χειρισμούς. Αυτή ακριβώς η αστάθεια δυσχεραίνει την προσπάθεια οργάνωσης της ροής των παραγωγικών μονάδων, και απαιτεί την άψογη εποπτεία της λειτουργίας του νοσοκομείου από το διοικητικό προσωπικό. Για να επιτευχθεί αυτό, οφείλει να υπάρχει δυνατότητα ενσωμάτωσης δεδομένων διαφορετικών πηγών που αυξάνονται συνεχώς, όπως τήρηση ιατρικών αρχείων, καταγραφή πληροφοριών ιατρικού προσωπικού, παρακολούθηση ασθενών σε πραγματικό χρόνο, εργαστηριακά πορίσματα κ.ό.κ.

Με αυτό τον τρόπο, η βελτίωση της αξιοποίησης των παραγωγικών μονάδων μπορεί να αυξήσει την ποιότητα και να μειώσει το κόστος της περίθαλψης.

4.4.8 Πρόληψη και έλεγχος των ενδονοσοκομειακών λοιμώξεων

Σύμφωνα με το Ευρωπαϊκό Κέντρο Πρόληψης και Ελέγχου Νοσημάτων (European Centre for Disease Prevention and Control), περίπου 100,000 ασθενείς εκτιμάται ότι προσβάλλονται από κάποια ενδονοσοκομειακή λοίμωξη κάθε έτος, εντός της Ευρωπαϊκής Ένωσης.⁵⁴ Ο αριθμός των θανάτων που καταγράφονται ως άμεση συνέπεια αυτών των λοιμώξεων υπολογίζεται ότι είναι τουλάχιστον 37,000 ενώ οι ίδιες λοιμώξεις θεωρείται ότι συμβάλλουν σε επιπλέον 110,000 θανάτους ετησίως. Εκτιμάται ότι περίπου έως και το 30% των νοσοκομειακών λοιμώξεων μπορούν να προληφθούν με κατάλληλα προγράμματα πρόληψης. Επιπλέον, σύμφωνα με τα Κέντρα Ελέγχου και Πρόληψης Ασθενειών στις ΗΠΑ, υπήρξαν περίπου 722,000 κρούσματα ενδονοσοκομειακών λοιμώξεων μόνο το 2011. Περίπου 75,000 ασθενείς έχασαν τη ζωή τους κατά τη διάρκεια της νοσηλείας τους λόγω αυτών.⁵⁵ Η πρόληψη των ενδονοσοκομειακών λοιμώξεων θα μπορούσε να εξοικονομήσει 25 ως 32 δισεκατομμύρια δολάρια ετησίως μόνο στις ΗΠΑ. Ο Παγκόσμιος Οργανισμός Υγείας (ΠΟΥ) έχει θέσει αυστηρές κατευθυντήριες γραμμές σχετικά με τα πρωτόκολλα που πρέπει να ακολουθούνται προκειμένου να ελαχιστοποιηθεί ο κίνδυνος της εξάπλωσης των λοιμώξεων εντός των νοσοκομείων. Ενώ μερικές από τις κατευθυντήριες γραμμές είναι εφικτό να εφαρμοσθούν και να τηρηθούν, υπάρχουν άλλες που είναι δύσκολο εξαιτίας της έλλειψης τεχνολογίας ικανής να διασφαλίσει την αυστηρή τήρηση των πρωτοκόλλων. Η αξιοποίηση των Big Data σε πραγματικό χρόνο είναι απαραίτητη ώστε να ενσωματωθούν γονιδιωματικά με επιδημιολογικά δεδομένα, να επιτευχθεί η απαραίτητη πρόληψη και να μειωθεί το μέγεθος του φαινομένου.

4.4.9 Δυνατότητα συμμετοχής του ασθενούς στη θεραπεία

Πέραν της συλλογής δεδομένων σχετικών με την αποτελεσματικότητα των μεθόδων θεραπείας των ασθενών, υπάρχουν και άλλες δυνατότητες συμμετοχής του ασθενούς στη φροντίδα. Χαρακτηριστικά παραδείγματα αποτελούν οι τρόποι φροντίδας που έχουν τον ασθενή στο επίκεντρο. Σε αυτές ο ίδιος ο ασθενής έχει γνώση των ιατρικών δεδομένων και παίρνει μέρος στη λήψη αποφάσεων που αφορούν τη θεραπεία του. Για να καταστεί αυτό δυνατό, απαιτείται η ενημέρωση του ασθενούς σχετικά με τη νόσο, τις επιλογές φαρμάκων, τα μειονεκτήματα και πλεονεκτήματα κάθε επιλογής, με τρόπο δομημένο και ικανό να επιτρέψει την επιλογή στρατηγικής με αντικειμενικά κριτήρια. Η χρήση των Big Data μπορεί να ενσωματώσει τα δεδομένα που αφορούν τον ασθενή και να του προσφέρει νέες δυνατότητες με αναβαθμισμένο ρόλο στη σύγχρονη Ιατρική επιστήμη.

4.4.10 Φροντίδα στο σπίτι

Ιδανικά, οι υπηρεσίες υγείας δεν θα έπρεπε να περιορίζονται εντός νοσοκομείων και κλινικών, αλλά να προσφέρονται ορισμένες δυνατότητες φροντίδας από το σπίτι. Πλέον υπάρχουν συσκευές που επιτρέπουν την παρακολούθηση του ασθενούς σε πραγματικό χρόνο μέσω επιλογών, όπως τα «έξυπνα» κινητά τηλέφωνα, ιατρικά όργανα που μπορεί να διαθέτει ο ασθενής (πιεσόμετρο, μετρητής σακχάρου κ.ό.κ.) και τα οποία η τεχνολογία των Big Data έχει επιτρέψει να διαμοιράζονται και αναλύονται αυτοματοποιημένα σε πραγματικό χρόνο. Με αυτόν τον τρόπο, η παραμονή του ασθενούς σε καθεστώς επίβλεψης, θα μπορούσε να περιοριστεί σε κάποιες περιπτώσεις και η παρακολούθηση να γίνεται εξ αποστάσεως. Το γεγονός αυτό πέραν του προφανούς οικονομικού αντίκτυπου, μπορεί να επηρεάσει την ψυχολογία των ασθενών, οι οποίοι συχνά επιζητούν το οικείο τους περιβάλλον, και εν τέλει την ίδια την υγεία τους, που αποδεδειγμένα συνδέεται με την ψυχολογική κατάσταση.

4.4.11 Συμβολή στην έρευνα

Η ενσωμάτωση και ανάλυση τεράστιου όγκου ιατρικών δεδομένων από διαφορετικές πηγές, όπως ηλεκτρονικά ιατρικά αρχεία, κοινωνικές εξελίξεις, επιστημονικά άρθρα και μελέτες, είναι δύσκολη υπόθεση για το ανθρώπινο δυναμικό που δραστηριοποιείται στην έρευνα στο χώρο της υγείας. Η χρήση των Big Data, μπορεί να συνδράμει στην Ιατρική και φαρμακευτική έρευνα, καθώς επιτρέπει:

- Ανακάλυψη κρυφών σχέσεων μεταξύ ετερογενών δεδομένων.
- Ανάπτυξη μοντέλων πρόγνωσης της εξέλιξης των ασθενειών και έλεγχος.
- Ανάλυση δεδομένων μεγάλων πληθυσμών, ώστε να ελέγχονται σε πραγματικό χρόνο πιθανά μοντέλα εξατομικευμένης καταλληλότητας φαρμάκων.

4.5 Ερευνητικό έργο και δράσεις

4.5.1 IBM Watson

Ο **IBM Watson** είναι ένας υπερυπολογιστής (supercomputer) τον οποίο κατασκεύασε η εταιρία IBM. Πρόκειται για ένα σύνθετο υπολογιστικό σύστημα, το οποίο είναι σε θέση να δίνει απαντήσεις σε ερωτήματα διατυπωμένα σε φυσική γλώσσα.

Αρχικά, δημιουργήθηκε με σκοπό να απαντάει τις ερωτήσεις ενός παιχνιδιού γνώσεων, με τίτλο Jeopardy. Το 2011 διαγωνίσθηκε όλους τους νικητές που είχαν συμμετάσχει στο παιχνίδι, επικρατώντας έναντι όλων και κερδίζοντας το βραβείο. Κατά τη διάρκεια των ερωτήσεων είχε πρόσβαση σε 200 εκατομμύρια σελίδες δομημένων και μη δομημένων δεδομένων, καταναλώνοντας 4 terabytes (10^{12} bytes) αποθηκευτικού χώρου στο δίσκο, συμπεριλαμβανομένων κειμένων της ιστοσελίδας Wikipedia, δίχως όμως να είναι συνδεδεμένος κατά τη διάρκεια του παιχνιδιού. Ήταν πιο αποτελεσματικός από όλους τους ανθρώπους που ανταγωνιζόταν, αλλά είχε σχετικά μεγαλύτερη δυσκολία όταν οι ερωτήσεις αποτελούνταν από λιγότερες λέξεις, καθώς η λειτουργία του στηρίζεται στην εύρεση λέξεων-κλειδιών και στον εντοπισμό συσχετίσεων μεταξύ αυτών.⁵⁶

Ο Watson IBM στην υγεία

Τον Φεβρουάριο του 2011, ανακοινώθηκε ότι η IBM θα συνεργάζεται με την εταιρία Nuance Communications στα πλαίσια ενός ερευνητικού προγράμματος με στόχο την ανάπτυξη ενός εμπορικού προϊόντος εκμετάλλευσης των δυνατοτήτων κλινικής υποστήριξης αποφάσεων του Watson, με προβλεπόμενη διάρκεια 18 έως 24 μήνες. Οι ιατροί στο Πανεπιστήμιο Columbia θα προσπαθούσαν να εντοπιστούν τα κρίσιμα ζητήματα στην άσκηση της Ιατρικής στα οποία η τεχνολογία του Watson είναι σε θέση να συμβάλλει, ενώ οι ιατροί στο Πανεπιστήμιο του Maryland θα επιχειρούσαν να προσδιορίσουν τον αποδοτικότερο τρόπο με τον οποίο η τεχνολογία του Watson θα μπορούσε να αλληλεπιδράσει με τους ιατρούς ούτως ώστε να παρέχεται η καλύτερη δυνατή φροντίδα στους ασθενείς.⁵⁷

Τον Σεπτέμβριο του 2011, οι εταιρίες IBM και WellPoint ανακοίνωσαν μια νέα συνεργασία ώστε να χρησιμοποιήσουν τα διαθέσιμα Big Data που διαθέτουν, στο σύστημα του Watson με στόχο την εύρεση νέων θεραπευτικών επιλογών για τους ιατρούς, οι οποίες δεν είχαν προταθεί μέχρι εκείνη τη στιγμή. Στη συνέχεια, το Φεβρουάριο του 2013, οι IBM και WellPoint ανακοίνωσαν ότι η πρώτη

εμπορική χρήση του Watson θα ήταν προσανατολισμένη στη λήψη αποφάσεων σχετικά με τη θεραπεία του καρκίνου των πνευμόνων. Ο Watson χρησιμοποιήθηκε για το σκοπό αυτό στην κλινική Memorial Sloan Kettering Cancer Center, στην Πολιτεία της Νέας Υόρκης.⁵⁸

Τον Οκτώβριο του 2012, η IBM ανακοίνωσε μία νέα συνεργασία με την κλινική του Cleveland. Συγκεκριμένα, η εταιρία έστειλε τον Watson στην κλινική Cleveland Clinic Lerner College of Medicine του Πανεπιστημίου Case Western Reserve, με σκοπό την αύξηση της τεχνογνωσίας του ιατρικού προσωπικού για τη θεραπεία των ασθενών. Το ανθρώπινο δυναμικό της κλινικής ανέλαβε την προσπάθεια αξιοποίησης της ικανότητας του Watson να αποθηκεύει και να επεξεργάζεται μεγάλες ποσότητες πληροφοριών, ενώ παράλληλα στόχευε στην επιτάχυνση και στην αύξηση της ακρίβειας της διαδικασίας επεξεργασίας. «Η συνεργασία της Cleveland Clinic με την IBM είναι συναρπαστική γιατί μας δίνει την ευκαιρία να διδάξουμε τον Watson να σκέφτεται με τρόπους που έχουν την προοπτική να τον μετατρέψουν σε ένα πανίσχυρο εργαλείο της Ιατρικής επιστήμης», δήλωσε ο C. Martin Harris, MD, επικεφαλής του τμήματος πληροφοριών της κλινικής του Cleveland.⁵⁹

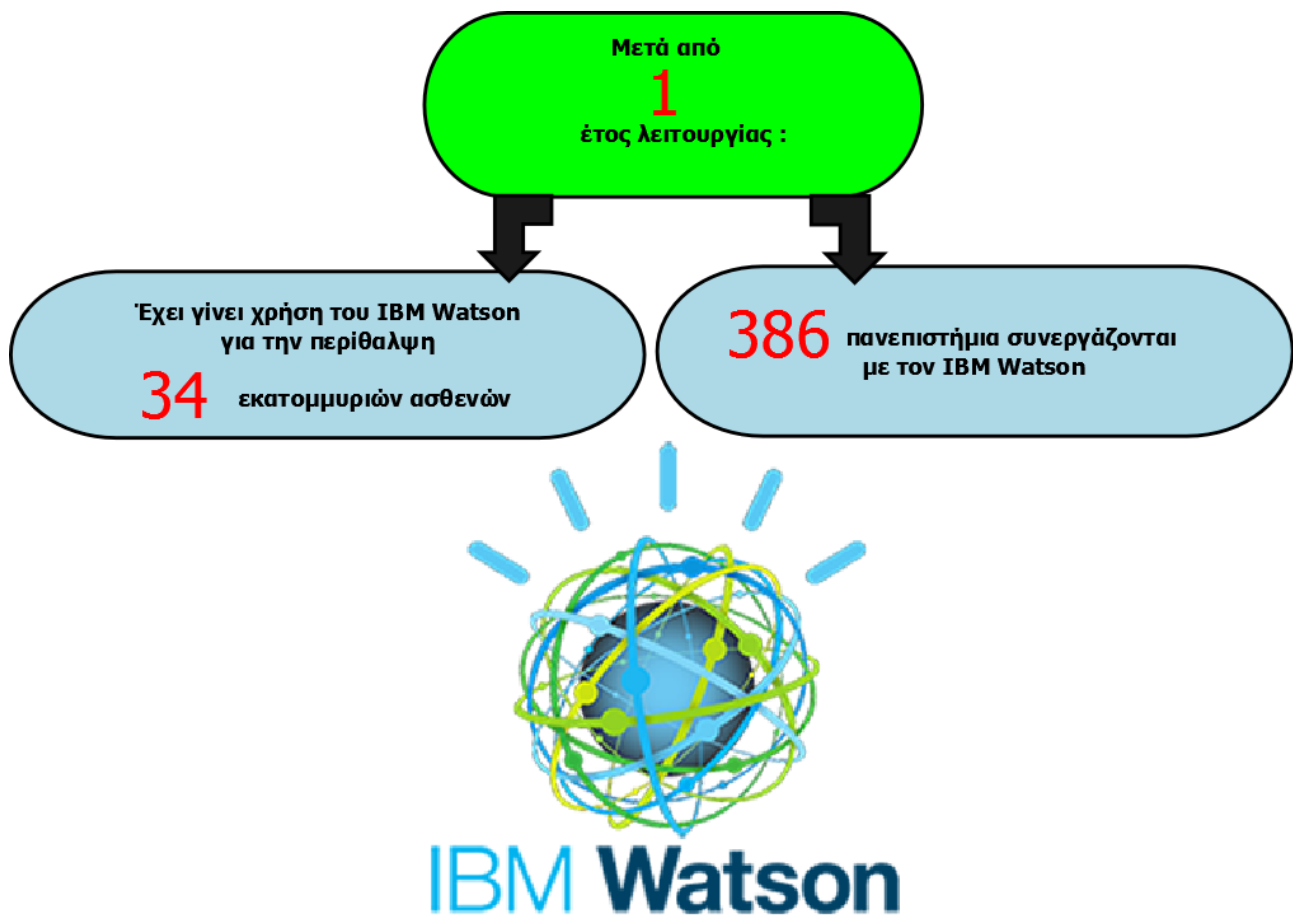
Το 2013, η IBM και η κλινική θεραπείας καρκίνου, MD Anderson Cancer Center, εγκαινίασαν από κοινού ένα πιλοτικό πρόγραμμα με σκοπό την εκπλήρωση του οράματος της κλινικής, δηλαδή «την εξάλειψη του καρκίνου». ^{60 61} Ωστόσο, αν και έχει λάβει επιχορήγηση 62 εκατομμυρίων δολλαρίων, το έργο δεν έχει επιτύχει τους στόχους του και τέθηκε προσωρινά σε αναμονή.⁶²

Στις 8 Φεβρουαρίου του 2013, η IBM ανακοίνωσε ότι Ογκολόγοι της κλινικής Maine Center for Cancer Medicine and WestMed Medical Group της Νέας Υόρκης έχουν αρχίσει δοκιμές του συστήματος του υπερυπολογιστή Watson στα πλαίσια της προσπάθειας για θεραπεία του καρκίνου του πνεύμονα.⁶³

Στις 29 Ιουλίου του 2016, η IBM και τα Manipal Hospitals (μια κορυφαία αλυσίδα νοσοκομείων στην Ινδία), ανακοίνωσαν την έναρξη της χρήσης του Watson στην Ογκολογία, σε ασθενείς με καρκίνο. Ο υπερυπολογιστής παρέχει πληροφορίες και ιδέες σε ιατρούς για τους ασθενείς με καρκίνο, ώστε να εντοπίσει εξατομικευμένες, βασισμένες σε στοιχεία, επιλογές θεραπείας του καρκίνου. Έτσι, η αλυσίδα Manipal Hospitals έγινε η πρώτη στον κόσμο που υιοθέτησε αυτή την τεχνολογία και η πρώτη που προσέφερε στους ασθενείς τη δυνατότητα πρόσβασης ώστε να λαμβάνουν μία δεύτερη γνώμη στην ιστοσελίδα τους μέσω του διαδικτύου.⁶⁴

Στις 7 Ιανουαρίου, το 2017, η IBM και η ασφαλιστική εταιρία Fukuoku Mutual Life Insurance συνάψαν σύμβαση. Σύμφωνα με αυτή, η IBM μέσω του Watson θα κάνει ανάλυση των αποζημιώσεων που παρέχει η ασφαλιστική εταιρία σε πελάτες. Τα πρώτα αποτελέσματα είναι αύξηση της παραγωγικότητας κατά 30% και εξοικονόμηση περίπου 1.5 εκατομμυρίων δολλαρίων στους πρώτους δύο μήνες λειτουργίας.⁶⁵

Ο Watson χαίρει πλέον της ευρείας αποδοχής του ιατρικού προσωπικού. Σύμφωνα με δηλώσεις του Manoj Saxena, πρώην επικεφαλής του τμήματος επιχειρήσεων υγείας της IBM, το 90% των νοσοκόμων που χρησιμοποιούν τον Watson, πλέον ακολουθούν πιστά τις προτάσεις του.⁶⁶



Εικόνα 4.8: Ο IBM Watson με αριθμούς

Τρόπος λειτουργίας

Το ερώτημα που γεννάται είναι ποιες λειτουργίες πραγματοποιεί ο Watson και με ποιο τρόπο καταφέρνει τα εκπληκτικά αυτά αποτελέσματα, αξιοποιώντας τα Big Data με τα οποία τροφοδοτείται. Η συνήθης διαδικασία που ακολουθείται είναι η ακόλουθη:

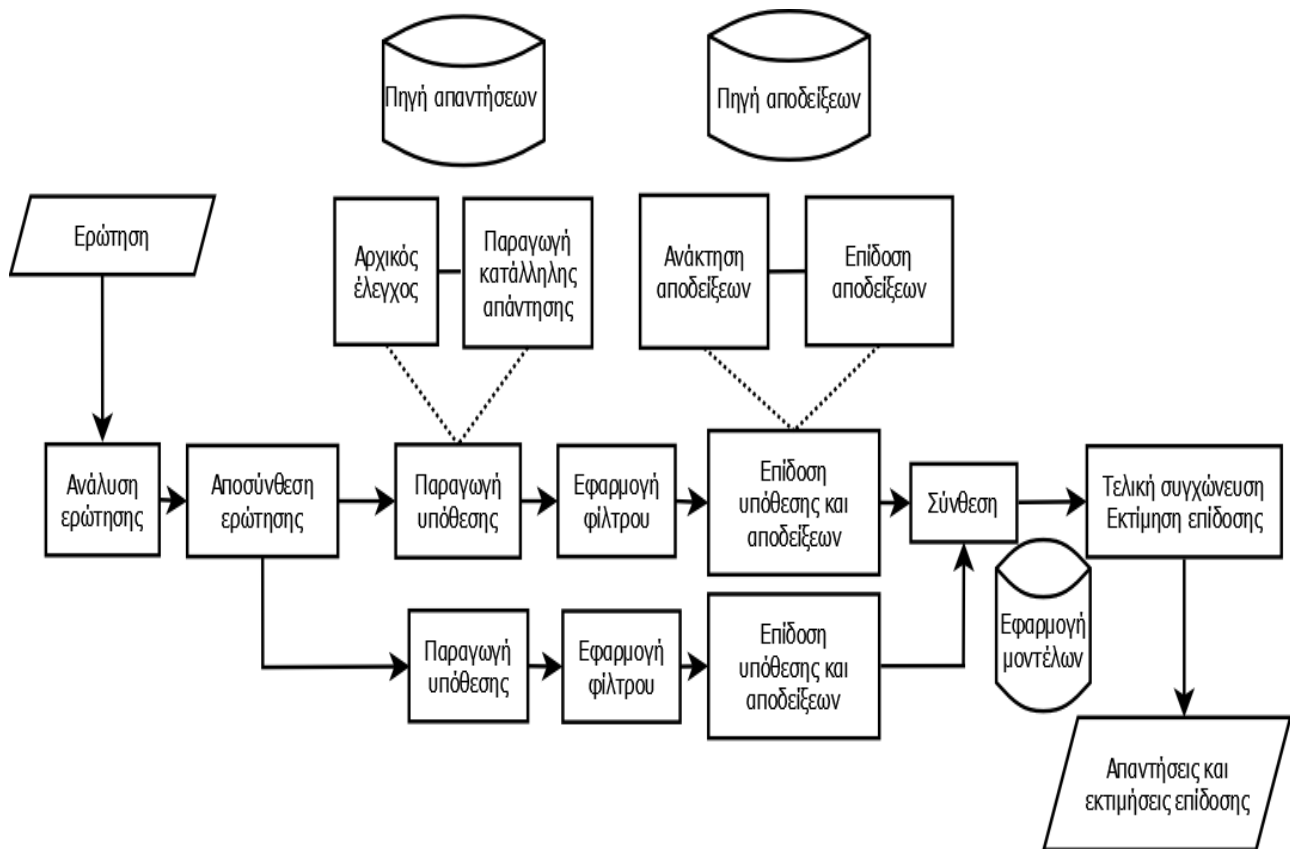
Το ιατρικό προσωπικό διατυπώνει σε φυσική γλώσσα το πρόβλημα που αντιμετωπίζει, περιγράφοντας τα συμπτώματα και άλλους σχετικούς παράγοντες.

Έπειτα, ο Watson εκτελεί τα εξής βήματα:

- ανατρέχει στα δεδομένα του ασθενούς
- διαχωρίζει τις λέξεις που ειπώθηκαν
- επιχειρεί να εκτιμήσει τις σημαντικότερες λέξεις-κλειδιά
- αναζητεί κοινά στοιχεία με το διαθέσιμο αρχείο ιατρικών δεδομένων
- σχηματίζει υποθέσεις
- διατυπώνει λίστα με τις πιθανές αιτίες δίνοντας μάλιστα αντίστοιχη βαρύτητα σε κάθε εκτίμηση

Οι πηγές των Big Data στις οποίες ανατρέχει ο Watson μπορεί να είναι προτεινόμενες μέθοδοι θεραπείας από τη βιβλιογραφία, σημειώσεις και καταγραφές ιατρών και νοσοκόμων, ηλεκτρονικά ιατρικά αρχεία ασθενών, κλινικές δοκιμές και έρευνες, επιστημονικά άρθρα, καθώς επίσης και πληροφορίες που παρέχονται από τους ίδιους τους ασθενείς. Αν και αναπτύχθηκε και διαφημίστηκε ως σύμβουλος διάγνωσης και θεραπείας, στην πραγματικότητα, ο Watson έχει χρησιμοποιηθεί κυρίως στην θεραπεία ασθενών που έχουν ήδη διαγνωσθεί με κάποια ασθένεια, προτείνοντας τρόπους αντιμετώπισης.

Η λειτουργία του Watson απεικονίζεται στο ακόλουθο διάγραμμα ροής της πληροφορίας:



Εικόνα 4.9: Η λειτουργία του IBM Watson σχηματικά

Παρουσιάζονται πέντε ενδεικτικοί τρόποι χρήσης Big Data Analytics από τον Watson IBM στην υπηρεσία της υγείας.⁶⁷

Ερμηνεία ιατρικών εικόνων

Το 2015, η IBM αγόρασε την εταιρία Merge Healthcare έναντι ενός δισεκατομμυρίου δολλαρίων. Η εξαγορά έδωσε πρόσβαση στην IBM σε περισσότερες από 30 δισεκατομμύρια ιατρικές εικόνες τα οποία προέρχονταν από 7.500 νοσοκομεία και κλινικές στις ΗΠΑ που ανήκαν στην Merge Healthcare. Αν και οι περισσότεροι άνθρωποι στον κόσμο δεν θα μπορούσαν να ωφεληθούν από τη χρήση των εικόνων που ανήκαν στον προσωπικό τους ιατρικό φάκελο, τα δεδομένα αυτά είναι ένα ορυχείο χρυσού για τον Watson.⁶⁸

Η IBM επιχείρησε να τροφοδοτήσει τον Watson με αυτές τις εικόνες με την ελπίδα ότι θα εντοπιστούν σημαντικές πληροφορίες τις οποίες οι ιατροί ήταν πιθανό να αγνοούν. Μέσω μάλιστα της

τεχνολογίας της μηχανικής μάθησης (machine learning), υπάρχει η δυνατότητα προσδιορισμού επαναλαμβανόμενων μοτίβων σε δεδομένα που αφορούν μεγάλες ποσότητες πληροφοριών.

Μέσω λοιπόν των επενδύσεων αυτών, η IBM κατέληξε πως ο Watson θα χρησιμοποιηθεί για τη μελέτη και διάγνωση μελανωμάτων, μια δύσκολα εντοπίσιμη μορφή καρκίνου του δέρματος, λόγω του τρόπου με τον οποίο εκδηλώνεται με διαφορετικό τρόπο σε κάθε άτομο. Χρησιμοποιώντας τις ικανότητες μηχανικής μάθησης του Watson, τέθηκε το ερώτημα και η πρόκληση, εάν το λογισμικό του θα καταφέρει αυτό που η Ιατρική επιστήμη δεν έχει καταφέρει σε μεγάλη κλίμακα.

Η σκέψη ήταν η εξής: οι ιατροί μπορούν να τροφοδοτήσουν μια εικόνα ενός ασθενούς που πρόσφατα διεγνώσθη με μελάνωμα, στο σύστημα του Watson. Έπειτα ο υπερυπολογιστής προσθέτει την εικόνα στις βάσεις δεδομένων του, έτσι ώστε να μπορεί να συγκριθεί με χιλιάδες άλλες εικόνες. Με τη συλλογή αυτών των πληροφοριών, οι ιατροί πλέον θα έχουν μια συνολική κατανόηση του πώς να διαγνώσουν στο μέλλον ασθενείς με αυξημένη πιθανότητα μελανώματος, καθώς και μια λίστα με τις επιλογές θεραπείας, η οποία προτείνεται από το σύστημα του Watson.

Ένα χρόνο μετά, οι επενδύσεις αυτές φαίνεται να αποδίδουν καρπούς. Οι ιατροί του Πανεπιστημίου της Βόρειας Καρολίνας εισήγαγαν στον Watson αρχεία εικόνων από 1000 ασθενείς με καρκίνο. Ο υπερυπολογιστής συμφώνησε με τις συστάσεις διακεκριμένων Ογκολόγων στο 99% των περιπτώσεων. Επιπλέον, ο Watson ήταν σε θέση να προτείνει πρόσθετες επιλογές από τους ιατρούς στο 30% των περιπτώσεων, αφού συγκέντρωσε όλες τις τελευταίες έρευνες για τον καρκίνο.

Θεραπεία σπάνιων μορφών παιδικών ασθενειών

Ένας στους δέκα κατοίκους των Η.Π.Α. πάσχουν από σπάνιες ασθένειες, σύμφωνα με μελέτες που προκύπτουν από το εγχείρημα Global Genes Project, οι μισοί εκ των οποίων είναι ανήλικοι. Έτσι, η IBM και η κλινική Boston Children's Hospital συνεργάστηκαν με την ελπίδα να χρησιμοποιηθούν οι δυνατότητες του Watson για να συμβάλλουν στον εντοπισμό, τη διάγνωση και τη θεραπεία σπάνιων παιδικών ασθενειών.⁶⁹

Αρχικά, η συνεργασία επικεντρώνεται στο ανθεκτικό σε στεροειδή νεφρωσικό σύνδρομο (steroid-resistant nephrotic syndrome ή SRNS), μια σπάνια γενετική μορφή νεφρικής νόσου που πλήττει παιδιά από την ηλικία των 2 ετών. Ο Watson καλείται να «διδασχθεί» νεφρολογία, την επιστήμη που

σχετίζεται με τη λειτουργία των νεφρών, διαβάζοντας την ιατρική βιβλιογραφία και αντλώντας περισσότερες πληροφορίες σχετικά με τις μεταλλάξεις που έχουν παρατηρηθεί στην ασθένεια αυτή, σε γονιδιακό επίπεδο. Οι ειδικοί από την παιδιατρική κλινική της Βοστώνης τροφοδότησαν τον Watson με Big Data που περιείχαν γενετικές πληροφορίες από τους προηγούμενους ασθενείς.

«Ένα από τα ταλέντα του Watson είναι η ταχύτατη εύρεση και σύνδεση μοτίβων σε σε μεγάλους όγκους μεγάλων δεδομένων», δήλωσε η Deborah DiSanzo, διευθύντρια του τμήματος υγείας του IBM Watson δελτίο Τύπου. Συμπλήρωσε πως «για τα παιδιά και τις οικογένειές τους που υποφέρουν χωρίς διάγνωση, στόχος μας είναι ο Watson να συνεργαστεί με τους κορυφαίους ειδικούς παγκοσμίως για να δημιουργήσει ένα γνωστικό εργαλείο που θα διευκολύνει τους ιατρούς να βρουν τη βελόνα στα άχυρα, αποκαλύπτοντας όλες τις σχετικές ιατρικές εξελίξεις και να υποστηρίξει αποτελεσματικά τη φροντίδα των παιδιών»⁶⁷

Μόλις ο Watson ολοκληρώσει το έργο του για το σύνδρομο SRNS, θα αναλυθεί η επιστημονική βιβλιογραφία και άλλες κλινικές βάσεις δεδομένων για να προσδιορισθούν γενετικές μεταλλάξεις σε περισσότερες ασθένειες και να προσδιορισθούν τρόποι ώστε οι ιατροί να εντοπίσουν τις κατάλληλες επιλογές θεραπείας. Ο σκοπός του έργου είναι να παραχθεί ένα σύστημα, το οποίο μπορεί να ερμηνεύσει αλληλουχίες του γονιδιώματος των παιδιών και όταν συνδυάζεται με την ιατρική βιβλιογραφία, θα δίνει τη δυνατότητα να προσδιορίζονται οι σπάνιες νόσοι πιο έγκαιρα και αποτελεσματικά από ποτέ.

Θεραπεία ασθενών με καρκίνο όταν υπάρχουν ελλείψεις στο ιατρικό προσωπικό

Σε μέρη όπου οι ασθενείς είναι πολύ περισσότεροι από τους ιατρούς, όπως στην Ινδία, η οποία διαθέτει μόλις και μετά βίας έναν ιατρό ανά 1.700 ασθενείς, η λήψη της κατάλληλης θεραπείας είναι ιδιαίτερα δύσκολη υπόθεση. Όμως, πρόσφατα, η IBM και η Manipal Hospital, το τρίτο μεγαλύτερο δίκτυο υγειονομικής περίθαλψης στην Ινδία, συνεργάστηκαν ώστε να χρησιμοποιήσουν τον Watson για τη διάγνωση και τη θεραπεία του καρκίνου σε 16 κλινικές και ακαδημαϊκά κέντρα, τα οποία αντιμετωπίζουν περίπου 200.000 ασθενείς με καρκίνο κάθε χρόνο.⁷⁰

Περισσότερα από ένα εκατομμύριο νέα κρούσματα καρκίνου διαγιγνώσκονται ετησίως στην Ινδία, όπου η έλλειψη σε ιατρούς είναι περίπου 16 φορές μεγαλύτερη από τις Η.Π.Α. «Είμαστε απελπιστικά ανεπαρκείς σε ιατρούς», δηλώνει ο Δρ Ajay Bakshi, διευθύνων σύμβουλος της Manipal Hospital, στο Forbes. «Ο Watson θα εξασφαλίσει την καλύτερη δυνατή φροντίδα των ασθενών».⁷⁰

Για το σκοπό αυτό σχεδιάστηκε ένα νέο λογισμικό, σε συνεργασία με το Memorial Sloan Kettering Cancer Center. Παρόμοια με τη χρήση του στο Boston Children's Hospital, ο Watson επεξεργάζεται Big Data που περιλαμβάνουν ιατρικές πληροφορίες από επιστημονικά περιοδικά και βιβλία και μετατρέπει την πληροφορία αυτή σε κατάλληλη μορφή ώστε είναι εύκολο κανείς να τη χρησιμοποιήσει για να ανακαλύψει προτάσεις σχετικές με τη θεραπεία του. Επίσης, το γεγονός ότι τροφοδοτείται συνεχώς η βάση δεδομένων του από νέες δημοσιεύσεις και νέα ιατρικά δεδομένα, συνεπάγεται ότι ο Watson, μέσω της μηχανικής μάθησης, θα βελτιώνεται συνεχώς παρέχοντας καλύτερα αποτελέσματα.

Όσο για το πώς ακριβώς ο Watson θα βοηθήσει τους ιατρούς στην Ινδία να αντιμετωπίσουν τον τεράστιο όγκο ασθενών, ο Δόκτωρ Bakshi, ανέφερε σε συνέντευξη τύπου πως «ο Watson θα βοηθήσει τους ιατρούς στην παροχή της πιο εξελιγμένης και αποτελεσματικής φροντίδας στους καρκινοπαθείς, με μικρό κόστος και σε ελάχιστο χρόνο».⁷¹

Κατανόηση γονιδιώματος του ασθενούς και παροχή εξατομικευμένων προτάσεων

Για τη μελέτη του ανθρώπινου γονιδιώματος σε ασθενείς με καρκίνο και την παροχή πρότασης θεραπείας ειδικά σχεδιασμένης για τον ασθενή, απαιτούνται συνήθως μερικές εβδομάδες, χρόνος που σε τέτοιες περιπτώσεις είναι πολύτιμος. Η IBM σε συνεργασία με 14 Ινστιτούτα που εξειδικεύονται στην καταπολέμηση του καρκίνου, υποστηρίζει πως ο Watson έχει τη δυνατότητα να προσφέρει τα αποτελέσματα αυτά σε λίγα μόλις λεπτά.⁷²

Μέσω των μαθησιακών ικανοτήτων του Watson, το DNA του ασθενούς θα μετατρέπεται σε ένα αξιοποιήσιμο γενετικό προφίλ, το οποίο συνδυαζόμενο με τη σχετική ιατρική βιβλιογραφία θα επιτρέψει στους ιατρούς να παρέχουν στους ασθενείς με τα πιο εξατομικευμένες επιλογές θεραπείας. Η μηχανική μάθηση θα συνεχίσει επίσης να βελτιώνει τις προτάσεις του Watson σχετικά με θεραπείες για τον καρκίνο, παρέχοντας λοιπόν τεράστια ποσότητα πληροφοριών σε συντομότερο χρονικό διάστημα.

Το γονιδίωμα ενός και μόνο ατόμου καταλαμβάνει περίπου 100 gigabytes μνήμης, δημιουργώντας μία μεγάλη συλλογή Big Data. Η ανάλυση των πληροφοριών αυτών από ιατρικούς φακέλους, επιστημονικά περιοδικά, σύγχρονες μελέτες και εγχειρίδια θα απαιτούσε αμέτρητες ώρες. Αντίθετα, οι ιατροί θα έχουν πλέον τη δυνατότητα να λάβουν όλη την πληροφορία σε λίγα λεπτά και να επικεντρωθούν στη θεραπεία των ασθενών τους. Ο Watson παρέχει τη δυνατότητα οπτικής απεικόνισης

των αποτελεσμάτων, μεθόδων θεραπείας και μεθόδων ελέγχου καταλληλότητας φαρμάκων, ειδικά για το συγκεκριμένο ασθενή.

«Όταν αντιμετωπίζουμε τον καρκίνο, δίνουμε ένα διαρκή αγώνα», δήλωσε ο Δόκτωρ Lukas Wartman, βοηθός διευθυντή στο τμήμα γονιδιωματικής έρευνας κατά του καρκίνου στο McDonnell Genome Institute στο Πανεπιστήμιο Ουάσιγκτον του Σεντ Λούις. «Όντας και εγώ καρκινοπαθής, καταλαβαίνω πόσο σημαντική είναι η παροχή γονιδιωματικής πληροφορίας. Δυστυχώς, η μετάφραση των γονιδιωμάτων που σχετίζονται με την εμφάνιση καρκίνου σε πιθανές θεραπευτικές επιλογές συχνά διαρκεί εβδομάδες και απαιτείται μια ομάδα εμπειρογνομόνων για τη μελέτη των όγκων για ένα μόνο ασθενή. Ο Watson είναι η μοναδική ρεαλιστική επιλογή ώστε να μειωθεί σημαντικά αυτό το χρονοδιάγραμμα».⁶⁶ Από την άποψη της τεχνολογίας, ο IBM Watson είναι ακόμα στα πρώτα του βήματα. Η τεχνολογία που θα επιτρέψει στον Watson να αξιοποιήσει στο έπακρο τις δυνατότητές του θα έρθει στο εγγύς μέλλον, αλλά η ουσία είναι ότι υπάρχει μέλλον. Η εξέλιξη οποιασδήποτε τεχνολογίας απαιτεί χρόνο. Αν και το έργο που συμβαίνει τώρα μπορεί να αυξήσει τις ελπίδες, ο Watson δεν μπορεί ακόμη να αντικαταστήσει τους ιατρούς, τους νοσηλευτές, ή τους άλλους επαγγελματίες υγειονομικής περίθαλψης. Ωστόσο, η ικανότητα του Watson να σαρώσει τα δισεκατομμύρια των ιατρικών εικόνων, να περιορίσει το χρόνο διάγνωσης, να παρέχει πληροφορίες για πιο εξατομικευμένη θεραπεία και να βοηθήσει το ιατρικό προσωπικό να διαχειριστεί το φορτίο, είναι μια συνεισφορά που ήδη του πιστοποιείται, ενώ το μέγεθος της συνεισφοράς είναι θέμα χρόνου να γίνει ορατό.

Ανάπτυξη εφαρμογών που προτείνουν συμβουλές διατροφής κατά την εγκυμοσύνη

Ενώ οι περισσότερες από τις προαναφερθείσες καινοτομίες θα βρεθούν σε νοσοκομεία και κέντρα έκτακτης ανάγκης σε όλο τον κόσμο, υπάρχει ένας τρόπος αξιοποίησης των Big Data, με τον οποίο ο Watson θα είναι διαθέσιμος σε κινητά τηλέφωνα απλών χρηστών.

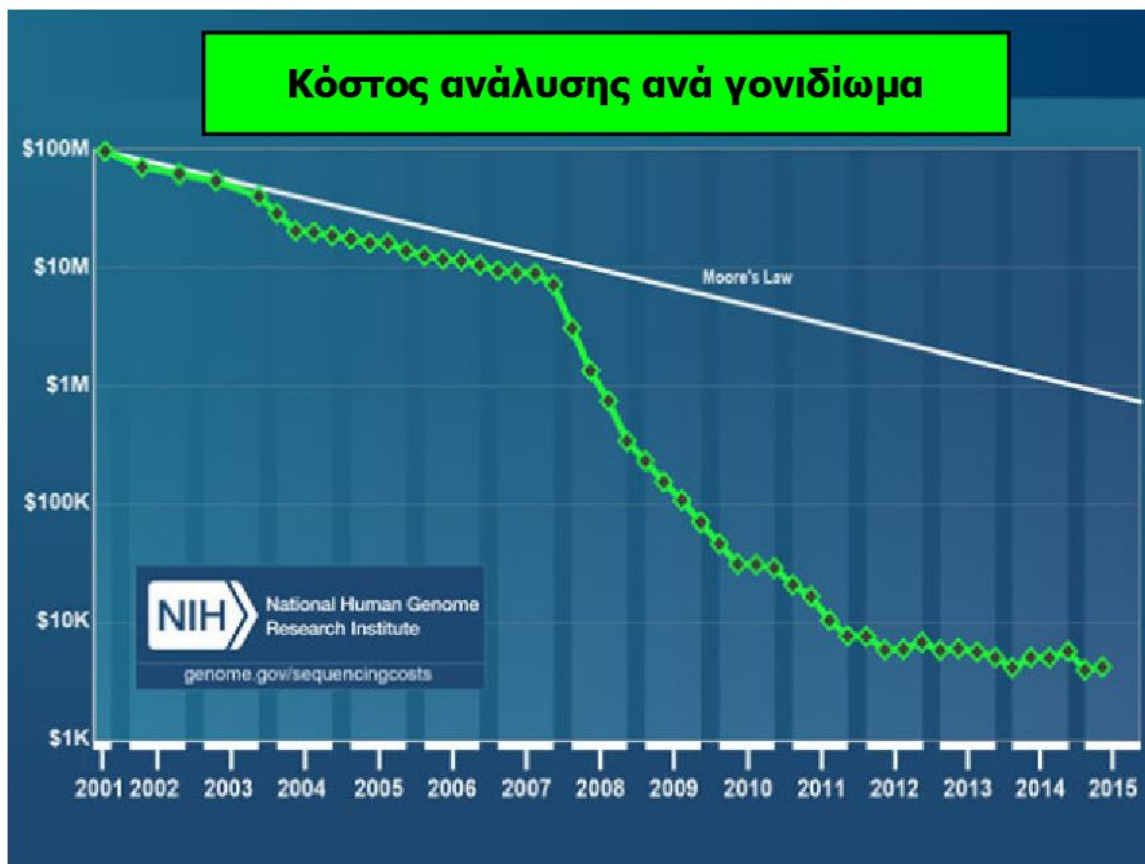
Η IBM παρουσίασε το Nutrino App Powered by Watson, τον Δεκέμβριο του 2015, μία συνεργασία με την εταιρία ανάπτυξης εφαρμογών υγιεινής διατροφής Nutrino Inc. Μέσω αυτής της εφαρμογής τα κινητά τηλέφωνα έχουν τη δυνατότητα να παρέχουν σε έγκυες γυναίκες «επιστημονικά τεκμηριωμένες και εξατομικευμένες συμβουλές διατροφής». Η εφαρμογή, μέσω χρήσης του Watson και των Big Data από τις βάσεις δεδομένων της Nutrino, προσφέρει εξατομικευμένες προτάσεις διατροφής και υποστήριξη καθ' όλη τη διάρκεια της ημέρας.⁷³

Μόλις μια έγκυος γυναίκα συνδεθεί στην εφαρμογή, μπορεί να εισάγει το στάδιο της εγκυμοσύνης, τις διατροφικές της προτιμήσεις και συνήθειες, τους στόχους για την υγεία, το σωματικό της βάρος, καθώς και τα δεδομένα που επιλέγονται από έξυπνες συσκευές που σχετίζονται με την υγεία. Έπειτα είναι δυνατόν να γίνουν ερωτήσεις σε φυσική γλώσσα προς τον Watson, σχετικά με την καταλληλότητα κάποιου γεύματος, ή για την παροχή πρότασης διατροφής για ολόκληρη την ημέρα. Ο Watson απαντά σε αυτά τα ερωτήματα μέσω αναζήτησης στη βάση δεδομένων της Nutrino, η οποία περιέχει πάνω από 500.000 τρόφιμα και 100.000 ιατρικές πηγές, σχηματίζοντας μία τεράστια συλλογή Big Data από δομημένα και μη δομημένα δεδομένα διαφορετικών πηγών.

4.5.2 Προτοβουλία εξατομικευμένης Ιατρικής

Οικονομικές επενδύσεις για το εγχείρημα

Η μελέτη του ανθρωπίνου γονιδιώματος είναι αδιαμφισβήτητα ένα σημαντικότερο εργαλείο για την Ιατρική επιστήμη. Τα τελευταία έτη η μείωση του χρηματικού κόστους ανάλυσης του γονιδιώματος, έχει ανοίξει το δρόμο σε ιδέες και καινοτομίες που στο παρελθόν φάνταζαν αδύνατες, όπως η εξατομικευμένη αντιμετώπιση των ασθενών. Στην ακόλουθη εικόνα γίνεται σαφής η εκθετική πτωτική πορεία που ακολουθεί το κόστος ανάλυσης ανά γονιδίωμα:



Εικόνα 4.10: Κόστος ανάλυσης ανά γονιδίωμα ως προς χρόνο (σε δολάρια)

Η πρωτοβουλία εξατομικευμένης Ιατρικής ξεκίνησε στις ΗΠΑ το 2015 και αποτέλεσε μία από τις σημαντικότερες προσπάθειες αναβάθμισης του συστήματος υγειονομικής περίθαλψης κατά τη διάρκεια της θητείας του Προέδρου Barack Obama. Μέσω αυτής επιχειρήθηκε η μετάβαση από τη φροντίδα προσαρμοσμένη στο «μέσο ασθενή» στη φροντίδα «προσαρμοσμένη στον κάθε ασθενή». Λαμβάνοντας υπόψη τα γονιδιώματα, το περιβάλλον και τον τρόπο ζωής του ασθενούς και αξιοποιώντας την τεχνολογία των Big Data, ο Watson έδωσε στους επαγγελματίες της υγείας τη δυνατότητα υιοθέτησης εξατομικευμένων μεθόδων θεραπείας.⁷⁴

Για τη στήριξη της δράσης αυτής διατέθηκαν 215 εκατομμύρια δολάρια μέσω συγχρηματοδότησης από το Εθνικό Ινστιτούτο Υγείας (**National Institute of Health** ή **NIH**) των ΗΠΑ, τον Οργανισμό Διαχείρισης Τροφίμων και Φαρμάκων (**Food and Drug Administration** ή **FDA**) και το Εθνικό Γραφείο Συντονισμού Τεχνολογίας της Πληροφορίας για την Υγεία (**Office of the National Coordinator for Health Information Technology** ή **ONCHIT**), εκ των οποίων:

- 130 εκατομμύρια δολάρια διατέθηκαν από το NIH για την ανάπτυξη εθελοντικής δράσης με τουλάχιστον ένα εκατομμύριο συμμετέχοντες, ώστε να εμπλουτιστεί η βάση με τα Big Data, να κατανοηθούν καλύτερα οι μηχανισμοί των ασθενειών και να πραγματοποιηθεί υπεύθυνος διαμοιρασμός των δεδομένων.
- 70 εκατομμύρια δολάρια διατέθηκαν από τον Εθνικό Ινστιτούτο Καρκίνου (National Cancer Institute ή NCI), τμήμα του NIH , ώστε να κλιμακωθούν οι προσπάθειες προσδιορισμού των γονιδιωματικών διαδικασιών δημιουργίας καρκίνου και εφαρμογής της γνώσης αυτής σε εξατομικευμένες προσπάθειες καταπολέμησης του καρκίνου.
- 10 εκατομμύρια δολάρια διατέθηκαν από τον FDA για την πρόσληψη επιστημονικού προσωπικού με στόχο τη δημιουργία υψηλής ποιότητας βάσεων με δεδομένα Big Data ώστε να είναι δυνατή η δυναμική διαχείρισή τους.
- 5 εκατομμύρια δολάρια διατέθηκαν από το ONC για τη στήριξη της διαλειτουργικότητας και της διαμόρφωσης πρωτοκόλλων και προδιαγραφών ασφαλούς και νόμιμης χρήσης των δεδομένων.

Στόχοι του Προγράμματος

Περισσότερες και αποτελεσματικότερες θεραπείες κατά του καρκίνου: Επιτάχυνση σχεδιασμού και ελέγχου αποτελεσματικών, εξατομικευμένων θεραπειών για την αντιμετώπιση του καρκίνου με την επέκταση των κλινικών δοκιμών σε γενετική βάση, διερεύνηση των θεμελιωδών βιολογικών πτυχών και την ίδρυση ενός «δικτύου γνώσης του καρκίνου» σε εθνικό επίπεδο για την τροφοδότηση της επιστημονικής έρευνας.

Δημιουργία εθελοντικής ερευνητικής ομάδας: Η ίδρυση μιας ομάδας αποτελούμενης από περισσότερους από 1 εκατομμύριο εθελοντές, η οποία σε συνεργασία με άλλους φορείς και παράγοντες, θα στηρίξει την πρωτοβουλία σε εθνική εμβέλεια. Οι εθελοντές θα πρέπει να συμμετέχουν στο σχεδιασμό της πρωτοβουλίας και να έχουν την ευκαιρία να συνεισφέρουν μέσω διαφορετικών πηγών Big Data, συμπεριλαμβανομένων των ιατρικών αρχείων γονιδιωματικών προφίλ, του προσδιορισμού των μικροοργανισμών του σώματος, του περιβάλλοντος και του τρόπου ζωής των ασθενών, δημιουργώντας έτσι πληθώρα αξιοποιήσιμων δεδομένων. Η προστασία των

προσωπικών δεδομένων θα πρέπει να τηρείται αυστηρά βάσει συγκεκριμένων προδιαγραφών. Το έργο αυτό στοχεύει στη στήριξη των ιατρικών ερευνών και στη δημιουργία κλινικού δίκτυου για την αξιοποίηση των καινοτόμων ερευνητικών μοντέλων που επιτρέπουν την ενεργή συμμετοχή των ασθενών. Τα δεδομένα της ομάδας θα είναι ευρέως προσβάσιμα σε εξειδικευμένους ερευνητές οι οποίοι θα έχουν τη δυνατότητα να εφαρμόσουν τη δημιουργική τους σκέψη για να δημιουργήσουν νέες ιδέες και προτάσεις θεραπείας. Τα πρότυπα διαλειτουργικότητας για τη διασφάλιση της ασφαλούς ανταλλαγής δεδομένων προϋποθέτουν τη συγκατάθεση των ασθενών.

Δέσμευση για την προστασία της ιδιωτικότητας: Για τη διασφάλιση της αρχής ότι η πρωτοβουλία θα συμμορφώνεται με τα νομικά πλαίσια προστασίας της ιδιωτικής ζωής, προβλέπεται η συμμετοχή διαφόρων ομοσπονδιακών υπηρεσιών ώστε τα στοιχεία που συλλέγονται από τις ομάδες ασθενών να συμμορφώνονται με τους κανόνες βιοηθικής και προστασία της ιδιωτικής ζωής και των ατομικών ελευθεριών.

Σύμπραξη ιδιωτικού και δημόσιου τομέα: Η πρωτοβουλία αποσκοπεί στην ενθάρρυνση συνεργασίας δημόσιων και ιδιωτικών φορέων μεταξύ τους καθώς και με ομάδες ασθενών ώστε να αναπτυχθούν οι απαραίτητες υποδομές για τη δημιουργία βάσεων δεδομένων με γονιδιώματα καρκινοπαθών. Για το λόγο αυτό συνεργάζονται ακαδημαϊκά ιατρικά και ερευνητικά κέντρα, κρατικά ιδρύματα, ιδιώτες, ιατροί και επαγγελματίες που δραστηριοποιούνται στον τομέα της υγείας ώστε η διακίνηση της γνώσης να συμβάλλει στους σκοπούς της πρωτοβουλίας.

4.5.3 Πλαίσιο Ευρωπαϊκής Ιατρικής Πληροφορίας - EMIF

Επενδύσεις και στόχοι του προγράμματος

Το Πλαίσιο Ευρωπαϊκής Ιατρικής Πληροφορίας, το οποίο αναφέρεται και ως **EMIF (European Medical Information Framework)**, δημιουργήθηκε τον Ιανουάριο του 2013, με σκοπό την αξιοποίηση των Big Data για τη βελτίωση της υγείας.⁷¹ Βασικός του στόχος είναι η δημιουργία ενός περιβάλλοντος το οποίο να επιτρέπει την αποδοτική επαναχρησιμοποίηση των διαθέσιμων ιατρικών δεδομένων, μέσω μίας πλατφόρμας, της **EMIF-Platform**. Χρηματοδοτείται από την Πρωτοβουλία Ιατρικής Καινοτομίας (**Innovative Medicines Initiative** ή **IMI**) Προκειμένου να διασφαλιστεί η άμεση εφαρμογή της δράσης, παράλληλα με την εκκίνηση του προγράμματος, ορίστηκαν δύο εξειδικευμένες ερευνητικές δράσεις με σκοπό να συμβάλλουν στην ανάπτυξη του Πλαισίου:

- **EMIF-AD:** Έρευνα που συμβάλλει στην αναγνώριση και επιβεβαίωση των παραγόντων που επισπεύδουν την εμφάνιση της νόσου Alzheimer (EMIF-Alzheimer Disease).
- **EMIF-Metabolic:** Έρευνα που στοχεύει στον εντοπισμό των μεταβολικών επιπλοκών της παχυσαρκίας.

Η κυριότερη λειτουργία της **EMIF-Platform** είναι η διαχείριση της επαναχρησιμοποίησης των ιατρικών Big Data. Δεδομένης της ποικιλίας των πηγών δεδομένων, τα οποία ενδέχεται να έχουν χρησιμότητα, η πλατφόρμα επιτρέπει την αναγνώριση, αξιολόγηση και επιλογή των κατάλληλων πηγών δεδομένων, από τον κατάλογο δεδομένων που διαθέτει (EMIF data catalogue). Επειδή τα δεδομένα προέρχονται από πολλές διαφορετικές πηγές και εντοπίζονται σε διαφορετικές μορφές και δομές, η πλατφόρμα δίνει έμφαση στη εναρμόνισή τους σύμφωνα με αυστηρά οριοθετημένες προδιαγραφές, ώστε να επιτρέπει την αποδοτική χρήση τους στην έρευνα. Γενικότερα, η ανάπτυξη της EMIF-Platform έγινε με στόχο να παρέχει δυνατότητες πρόσβασης στα ιατρικά δεδομένα, ανάλυσης και οπτικής απεικόνισης. Τα προγράμματα EMIF-AD και EMIF-Metabolic χρησιμοποιήθηκαν ως σενάρια χρήσης και τα αντίστοιχα αποτελέσματα διαμόρφωσαν τον τρόπο με τον οποίο πραγματοποιούνται οι επιθυμητές λειτουργίες μέσω της πλατφόρμας.

Ο βασικός στόχος του ερευνητικού προγράμματος EMIF-AD είναι η βελτίωση του σχεδιασμού των μελετών θεραπείας και πρόληψης για τη νόσο του Alzheimer (AD) σε άτομα που δεν έχουν νοσήσει από τη νόσο. Για να το πετύχει αυτό, έχει τεθεί ως στόχος η ανακάλυψη και επικύρωση διαγνωστικών δεικτών, προγνωστικών δεικτών και παραγόντων κινδύνου για την εμφάνιση της AD σε υγιή άτομα. Το EMIF-AD κάνει χρήση τόσο των υφιστάμενων όσο και των νέων δεδομένων που συλλέγονται. Βασικοί στόχοι της EMIF-AD είναι η ανάπτυξη μιας πλατφόρμας δεδομένων σε συνεργασία με την EMIF-Platform. Η πλατφόρμα υποστηρίζει την πρόσβαση, αποθήκευση, συγκέντρωση και ανάλυση δεδομένων σε μεγάλες ομάδες. Για τους νοσούντες εφαρμόζεται διαφορετική μεθοδολογία για την ανακάλυψη νέων δεικτών. Οι υποψήφιοι δείκτες θα επικυρωθούν μέσω ελέγχου σε μεγάλα σύνολα δεδομένων πληθυσμού διαθέσιμα μέσω της EMIF-Platform. Οι δείκτες μπορεί να είναι βιοδείκτες (όπως τα ποσοστά της τοξικής πρωτεΐνης βήτα αμυλοειδούς που θεωρείται ότι είναι υπεύθυνη για την εμφάνιση της νόσου), γνωστικοί δείκτες (όπως ρυθμός μείωσης γνωστικών ικανοτήτων), ή κλινικοί (όπως ανθεκτικότητα στην άνοια σε μεγαλύτερης ηλικίας πληθυσμιακές ομάδες).

Το EMIF-Metabolic επιχειρεί την ταυτοποίηση νέων βιολογικών δεικτών και μηχανισμών για μεταβολικές επιπλοκές που σχετίζονται με την εμφάνιση παχυσαρκίας εξετάζοντας παθογόνους φαινοτύπους. Το πρόγραμμα διερευνά την ετερογένεια στις μεταβολικές συνέπειες της παχυσαρκίας

σε μικρούς, μεσαίους και μεγάλους πληθυσμούς. Με βάση αυτό, παράγεται μια περιγραφική βάση δεδομένων με συντελεστές κινδύνου εμφάνισης παχυσαρκίας βασισμένων στους φαινότυπους. Το EMIF-Metabolic παράγει δεδομένα παρατηρήσεων και δεδομένα από τυχαιοποιημένες ελεγχόμενες μελέτες για την ταξινόμηση των δεικτών παχυσαρκίας και των ασθενειών που σχετίζονται με αυτή. Προϊόν της μελέτης είναι βιοδείκτες σε πραγματικό περιβάλλον, καθώς και επιλογές πρόληψης.

Αποτίμηση του προγράμματος

Έπειτα από τρία έτη λειτουργίας, πραγματοποιήθηκε αξιολόγηση της EMIF-Platform και εκδόθηκε η τρίτη έκδοση καταλόγου (EMIF Catalogue v3). Η λειτουργία των EMIF-AD και EMIF-Metabolic αποδείχθηκε καθοριστική για την αποτίμηση του καταλόγου και τη διαμόρφωσή του, ενώ προστέθηκαν νέες δοκιμές ελέγχου ώστε να απαντηθούν ερευνητικές ερωτήσεις μέσω της εξόρυξης των κατάλληλων δεδομένων. Το αρχικό λογισμικό έχει αναβαθμιστεί με νέες προσθήκες και εργαλεία βελτιστοποίησης της ροής δεδομένων, δημιουργώντας ένα εργαλείο διαχείρισης (TASKA). Για την αποτίμηση της EMIF-Platform χρησιμοποιήθηκε το **OMOP Common Model (Observational Medical Outcomes Partnership)**.⁷⁶

Το EMIF-AD έχει αποδώσει τα αναμενόμενα, βάσει της αποτίμησης. Όσον αφορά τις υποδομές, έχουν βελτιωθεί σημαντικά η λειτουργικότητα και το περιεχόμενο του καταλόγου EMIF. Στα νέα δεδομένα που δημιουργήθηκαν περιλαμβάνονται 124 πρότυποι φαινότυποι ασθενών, σε σχέση με συγκεκριμένα γονίδια που σχετίζονται με τη νόσο Alzheimer. Πραγματοποιήθηκε ανάλυση μεγάλου αριθμού συνόλων δεδομένων που οδήγησε σε σημαντικές δημοσιεύσεις σχετικά με την έκβαση της άνοιας, το ποσοστό της τοξικής πρωτεΐνης βήτα αμυλοειδούς και της προδιάθεσης για τη νόσο. Επιπλέον, πραγματοποιήθηκαν μελέτες βιοδεικτών που οδήγησαν στην ανακάλυψη παθογενών φαινότυπων RNA και στην απεικόνισή τους.

Το EMIF-Metabolic επίσης εξελίχθηκε όπως έχει προγραμματιστεί με μικρές μόνο αποκλίσεις. Μελετήθηκε ο ρόλος της αντίστασης στην ινσουλίνη (insulin resistance) ως κοινή παθολογική εκδήλωση για την ανάπτυξη παχυσαρκίας και διαβήτη τύπου 2. Εκτεταμένες μεταβολικές μελέτες διεξήχθησαν σε προσεκτικά επιλεγμένα άτομα με παθήσεις του ήπατος και φαινότυπους που σχετίζονται με το διαβήτη. Τα στοιχεία έχουν ενσωματωθεί και αναλυθεί σε μια προσέγγιση βιολογικών συστημάτων, που οδήγησε στην ταυτοποίηση των πιθανών θεραπευτικών επιλογών και σε μια πιλοτική κλινική μελέτη. Η επικύρωση των βιοδεικτών και πρόσθετες μελέτες σε μεγάλες ομάδες βρίσκονται σε εξέλιξη. Πιθανές αιτίες της παχυσαρκίας και της νόσου του Alzheimer, καθώς και του καρκίνου του ενδομητρίου έχουν αναλυθεί χρησιμοποιώντας μια προσέγγιση τυχαιοποίησης

στα πρότυπα της διαδικασίας Medel οι οποίες έχουν ήδη δημοσιευθεί. Μια ανασκόπηση των γνωστών βιοδεικτών για διαβήτη τύπου 2 έχει πραγματοποιηθεί και δημοσιευθεί, η οποία επισημαίνει ότι οι πληροφορίες σχετικά με την αιτιότητα είναι σε μεγάλο βαθμό απύσες. Μετά από μελέτη διαταραχών παχυσαρκίας που σχετίζονται με διαβήτη τύπου 2, μη αλκοολική λιπώδη ηπατική νόσο, και κίρρωση του ήπατος σε μεγάλες βάσεις δεδομένων, προέκυψαν συμπεράσματα που χρησιμοποιήθηκαν σε πολλές ερευνητικές μελέτες. Τεράστιο επίτευγμα αποτελεί η δημιουργία, αποδοχή και διάδοση ενός πρωτοκόλλου με προδιαγραφές για την εξέταση της υγείας του ήπατος μέσω βιοδεικτών και των παραγόντων που την επηρεάζουν.

Επιτεύγματα από την αρχή του προγράμματος:

- Πρόοδος στην ανάπτυξη της αρχιτεκτονικής σχεδίασης της πλατφόρμας
- Τρίτη έκδοση καταλόγου
- Προσθήκη δυνατότητας εξωτερικής συμμετοχής και δημιουργίας ατομικών υποκαταλόγων
- Επιπλέον μελέτη νόμων και κώδικα βιοηθικής για το διαμοιρασμό των δεδομένων
- Βελτίωση ροής δεδομένων για την υποστήριξη της εξόρυξής τους με το εργαλείο TASKA
- Χρήση του OMOP Common Data Model για έλεγχο και προσθήκη νέων εργαλείων
- Δημιουργία πλήρους πρωτοκόλλου προδιαγραφών για μελέτη ασθενειών:
 - Νόσος Alzheimer: Σχεδίαση ελέγχου κατά περίπτωση, φλεγμονές και άνοια, μέθοδοι θεραπείας και πρόληψης
 - Μεταβολικές διαταραχές: Καρδιαγγιακές παθήσεις και δείκτης μάζας σώματος, παθήσεις του ήπατος
 - Έναρξη νέων εξειδικευμένων μελετών για δημιουργία πρωτοκόλλων
- Βελτιώσεις στο λογισμικό (Jerboa Reloaded) και νέες προσθήκες για πιο σύνθετες περιπτώσεις
- Ορισμός στρατηγικού πλάνου για εκτίμηση αξίας των δεδομένων
- Προσθήκη νέων πηγών δεδομένων από συνεργασίες για την αύξηση του όγκου των δεδομένων
- 85 νέες δημοσιεύσεις και επιστημονικές μελέτες βασισμένες στα αποτελέσματα της πλατφόρμας⁷⁷

4.5.4 OpenPHACTS

Στόχος του προγράμματος

Το πρόγραμμα OpenPHACTS (**Open Pharmacological Concept Triple Store**) εγκαινιάστηκε το 2013 και αποτελεί μία Ευρωπαϊκή πρωτοβουλία σύμπραξης δημόσιου και ιδιωτικού τομέα μεταξύ ερευνητών, ακαδημαϊκού χώρου, επιχειρήσεων, φαρμακοβιομηχανιών και άλλων οργανισμών με στόχο την αποδοτικότερη, οικονομικότερη και ταχύτερη ανακάλυψη νέων φαρμάκων.⁷⁸ Χρηματοδοτείται από τον οργανισμό IMI, όπως και το πλαίσιο EMIF, στα πλαίσια της σχεδίασης μεθόδων για τη δημιουργία υποδομών και προδιαγραφών κοινής χρήσης των δεδομένων για αποδοτικότερη ανάπτυξη φαρμάκων και φροντίδα ασθενών στο μέλλον.^{79 80}

Στα πλαίσια του προγράμματος υπάρχει συνεργασία με 27 Πανεπιστήμια σε όλη την Ευρώπη, 6 φαρμακευτικές εταιρίες και 4 εταιρίες που δραστηριοποιούνται στο χώρο των Big Data. Βασικός σκοπός είναι η εξάλειψη των εμποδίων για τη φαρμακευτική έρευνα, ενώ τα αποτελέσματα της χρήσης είναι ελεύθερα και διαθέσιμα στην πλατφόρμα GitHub.⁸¹ Η αξιοποίηση των δεδομένων γίνεται μέσω της εφαρμογής **Open PHACTS Discovery Platform**, η οποία διατίθεται δωρεάν και ενσωματώνει φαρμακευτικά δεδομένα από μεγάλο πλήθος πηγών, ενώ παρέχει εργαλεία και υπηρεσίες εξόρυξης δεδομένων, μέσω μίας φιλικής προς το χρήστη διεπαφής.⁸²

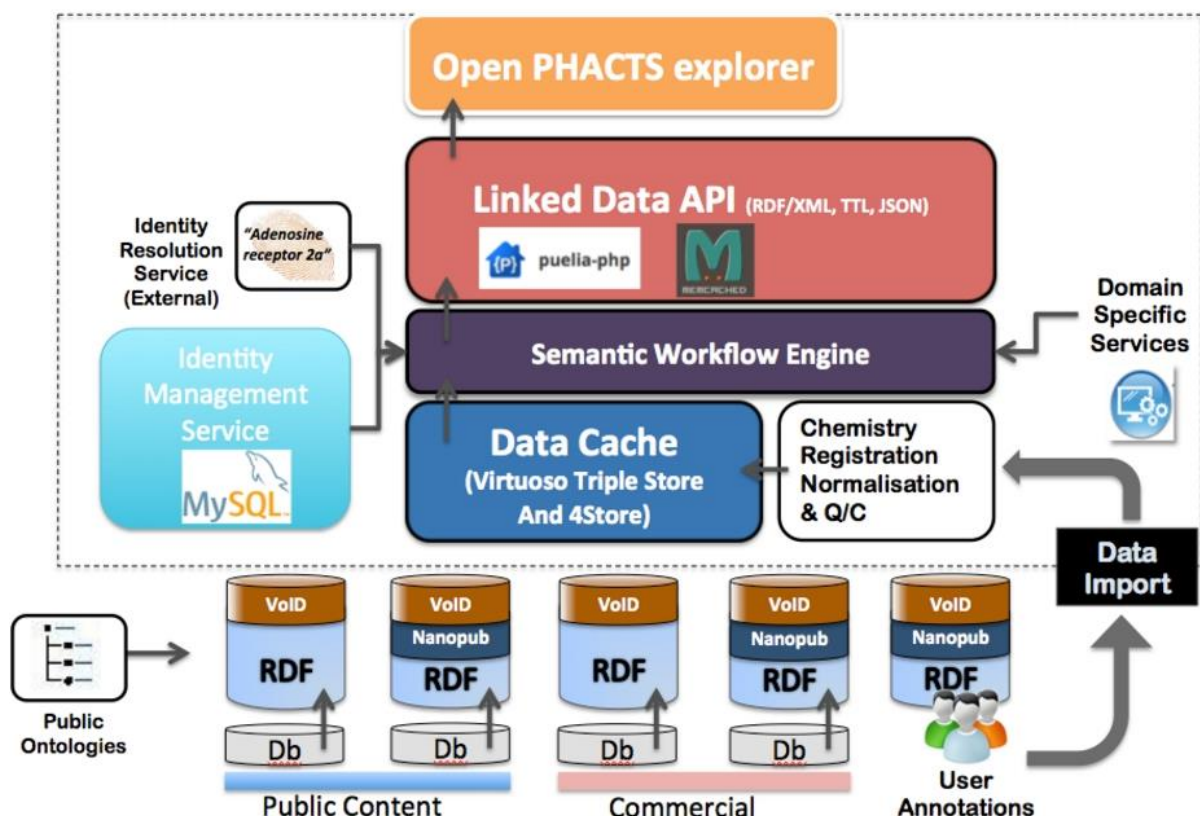
Τρόπος λειτουργίας

Ο μεγάλος αριθμός των διαθέσιμων βάσεων δεδομένων στον τομέα της ανάπτυξης φαρμάκων δημιουργεί την ανάγκη προσδιορισμού προτεραιοτήτων και μεθόδων για την επιλογή των κατάλληλων πληροφοριών σε ένα τεράστιο σύμπαν από Big Data . Στο πλαίσιο η πρωτοβουλία Open PHACTS υλοποιεί την αναζήτηση βάσει της σημασιολογικής βαρύτητας των ερευνητικών ερωτημάτων που πραγματοποιούνται στα πλαίσια της φαρμακευτικής έρευνας. Ο κατάλογος με τα συνηθέστερα και σημαντικότερα ερωτήματα ξεκίνησε από επιστήμονες ευρωπαϊκών φαρμακευτικών εταιρειών και στη συνέχεια η λίστα επεκτάθηκε και βελτιώθηκε συμπεριλαμβάνοντας ακαδημαϊκές έρευνες. Αρχικά, ορίστηκε ένα σύνολο 83 ερωτήσεων, οι οποίες ήταν ομαδοποιημένες κατά τομέα και κατά προτεραιότητα. Όλες οι ερωτήσεις γίνονται σε φυσική γλώσσα και απαιτούν την ενσωμάτωση τουλάχιστον δύο διαφορετικών πηγών δεδομένων.

Η πλατφόρμα Open PHACTS αποτελείται από τα ακόλουθα δομικά στοιχεία:

- **Πηγές δεδομένων (Data Sources):** Οι βάσεις δεδομένων περιέχουν πληροφορίες από πολλούς κλάδους, όπως χημεία, φαρμακολογία, δεδομένα ασθενειών, γονιδιωματικές βιβλιοθήκες, πρωτεϊνικές αλυσίδες κ.ό.κ.
- **Κρυφή Μνήμη Συνδεδεμένων Δεδομένων (Linked Data Cache):** Τα δεδομένα είναι αποθηκευμένα σε ένα κεντρικό αποθετήριο από όπου και γίνεται η εξόρυξή τους, για λόγους αξιοπιστίας και απόδοσης.
- **Υπηρεσία Ανάλυσης Ταυτότητας (Identity Resolution Service):** Πραγματοποιεί τη λειτουργία αντιστοίχισης των λέξεων που εισάγει ο χρήστης με τη μορφή ερωτημάτων σε σημασιολογικές έννοιες, γνωστές στο σύστημα.
- **Υπηρεσία Χαρτογράφησης Ταυτότητας (Identity Mapping Service):** Εφαρμόζει τη λειτουργία σύνδεσης των ερωτημάτων με σύνολα δεδομένων βάσει των συμφραζομένων. Ουσιαστικά κάνει εκτίμηση για το είδος της πληροφορίας που αναζητεί ο χρήστης εξαιρώντας σημασιολογικές έννοιες που προκύπτουν από την υπηρεσία IRS και θεωρούνται περιττές για τη συγκεκριμένη αναζήτηση και έχουν προκύψει από κοινή ορολογία για διαφορετικές έννοιες.
- **Ειδικές Υπηρεσίες Κατόχου (Domain Specific Services):** Υπάρχει πληθώρα σημαντικών φαρμακευτικών λειτουργιών ανάλογα με τον κάτοχο. Η ορθή επιλογή αυτών επηρεάζει την επίδοση του συστήματος. Ένα παράδειγμα αποτελεί η χαρτογράφηση στοιχείων βάσει χημικών δομών, αντί ονομάτων. Αντί αυτή να γίνει εξ ολοκλήρου από την αρχή, είναι δυνατόν να βασιστούμε στην υπάρχουσα χαρτογράφηση, εκτελώντας τις κατάλληλες τροποποιήσεις.
- **Πυρήνας API (Core Application Programming Interface):** Παρέχει ένα σύνολο βιβλιοθηκών και μεθόδων τις οποίες μπορούν να χρησιμοποιήσουν οι εφαρμογές. Με τον τρόπο αυτό, διευκολύνονται οι προγραμματιστές, καθώς τους παρέχονται κάποια ολοκληρωμένα εργαλεία για την αξιοποίηση των δεδομένων.

Ακολουθεί το λειτουργικό διάγραμμα της πλατφόρμας:



Εικόνα 4.11: λειτουργικό διάγραμμα της πλατφόρμας Open Phacts

Αντίκτυπος

Το πρόγραμμα Open PHACTS έχει σαφή αντίκτυπο με διάφορους τρόπους. Η πιο σημαντική συνεισφορά είναι η χρήση του συστήματος στην επιστημονική έρευνα. Αρκετές επιστημονικές δημοσιεύσεις προέρχονται από την εκτεταμένη χρήση του συστήματος, το οποίο επιτρέπει την ανάλυση δεδομένων που ήταν πολύ δύσκολο να επιτευχθεί στο παρελθόν. Πολλές φαρμακευτικές εταιρείες έχουν ενσωματώσει τα εσωτερικά τους δεδομένα μέσω του Open PHACTS, ώστε να μπορούν εύκολα να πραγματοποιήσουν ερωτήματα σε όλες τις πληροφορίες που είναι στη διάθεσή τους, τόσο στις δημόσιες όσο και στις ιδιωτικές.

Μια ακόμα συνεισφορά προέρχεται από την διαπίστωση ότι μεγάλες ποσότητες ποικίλων σημασιολογικών φαρμακευτικών δεδομένων μπορούν να αναλυθούν με αποδοτικό τρόπο, κάτι το

οποίο επιβεβαιώνεται από σημαντικούς φορείς, όπως το Ευρωπαϊκό Ινστιτούτο Βιοπληροφορικής (European Bioinformatics Institute ή EBI) και εμπορικούς παρόχους όπως η Thomson-Reuters. Η επιτυχία του προγράμματος Open PHACTS έχει αποδείξει την πρακτικότητα της χρήσης των Data στη βιοϊατρική έρευνα. Μάλιστα, το γεγονός ότι οι πάροχοι επέλεξαν να προσφέρουν τα δεδομένα τους, ενισχύει την αξία της δράσης και βοηθά τα μέγιστα στην διατήρηση του συστήματος Open PHACTS.

Τα αποτελέσματα του καταλόγου Open PHACTS έχουν επίσης τροφοδοτήσει συζητήσεις σχετικά με τη σημασία της διαλειτουργικότητας των δεδομένων που δημιουργούνται στην Ευρωπαϊκή Ένωση υπό την αιγίδα του IMI, καθώς και δημόσιων και ιδιωτικών φορέων. Αναγνωρίζεται ότι μια προσέγγιση όπως αυτή που υιοθετείται από την Open PHACTS πρέπει να τεθεί σε εφαρμογή για την παραγωγή των κατάλληλων φαρμάκων και την υποστήριξη των υπηρεσιών υγείας. Ενδεικτικό για τη χρησιμότητα της πλατφόρμας είναι το γεγονός ότι με τη συμπλήρωση δύο ετών από τη λειτουργία της, αριθμούσε ήδη περισσότερα από 500 εκατομμύρια ερωτήματα.

4.6 Ιστορίες επιτυχίας στη φροντίδα ασθενών

4.6.1 Η τεχνολογία των Big Data συνέβαλε στη μερική ανάκτηση της κίνησης παράλυτου ασθενούς

Η τεχνολογία των Big Data, έχει τη δυνατότητα να συνεισφέρει με διάφορους τρόπους στη φροντίδα της υγείας, πέραν της δημιουργίας πρωτοκόλλων λειτουργίας. Χαρακτηριστικό παράδειγμα είναι η ανάκτηση της ικανότητας κίνησης από παράλυτο ασθενή.⁸³

Ο τεράστιος όγκος των δεδομένων που είναι δυνατόν να παραχθεί από τη μελέτη της προσπάθειας κίνησης των άκρων, σε ένα παράλυτο άτομο, συγκριτικά με την κίνηση ενός υγιούς ατόμου, είναι υπερπολύτιμος. Συγκεκριμένα, μπορεί να δώσει πληροφορίες που θα επιτρέψουν ορισμένες μηχανικές κινήσεις. Η προοπτική αυτή βρίσκεται ακόμα σε πειραματικό στάδιο, ενώ το μέγεθος των συσκευών που χρησιμοποιούνται για να δώσουν αυτές τις δυνατότητες είναι ακόμα μεγάλο για να χρησιμοποιηθούν εκτός του ελεγχόμενου ερευνητικού περιβάλλοντος. Σε κάθε περίπτωση αυτή η χρήση των Big Data, είναι άξια αναφοράς, καθώς προσφέρει ελπίδες για την πλήρη επαναφορά της λειτουργικότητας των άκρων σε εκατομμύρια ασθενείς ανά τον κόσμο.

Η τεχνολογία αυτή αναπτύσσεται από την εταιρία Battelle, έναν μη-κερδοσκοπικό ερευνητικό οργανισμό, σε συνεργασία με το Ohio State University Wexner Medical Center. Η λειτουργία της βασίζεται στην παράκαμψη των νεύρων και τη μετατροπή της σκέψης των ασθενών σε σήματα που διαπερνούν τα μη-λειτουργικά τμήματα του νωτιαίου μυελού. Μέσω των Big Data Analytics, μελετώνται δεδομένα από 3 εκατομμύρια σημεία του εγκεφάλου του ασθενούς κάθε δευτερόλεπτο.

Η μελέτη αυτή αποτελεί μέρος ενός μεγαλύτερου προγράμματος χρηματοδοτούμενου από τον Οργανισμό Διαχείρισης Τροφίμων και Φαρμάκων (Food and Drug Administration). Ξεκίνησε το 2014, όταν ο Ian Burkhardt, πάσχον τετραπληγίας, έπειτα από ένα ατύχημα που είχε, έδωσε τη συγκατάθεσή του ώστε να του εμφυτευθεί στον εγκέφαλο μέσω εγχείρησης μία ηλεκτρονική πλακέτα, η οποία θα επέτρεπε στους ερευνητές να καταγράφουν την εγκεφαλική του δραστηριότητα.

Η εταιρία Battelle, έχει επικεντρώσει το έργο της στην ανάπτυξη αλγορίθμων οι οποίοι ταξινομούν τις τεράστιες ποσότητες ετερογενών Big Data που καταγράφονται, ώστε να εντοπιστούν τα σημεία και τα νεύρα που είναι τα σημαντικότερα για την κίνηση. Με τη σκέψη τα σήματα που παράγονται μετατρέπονται σε πληροφορία αξιοποιήσιμη από υπολογιστικά συστήματα, μέσω 160 ηλεκτροδίων

που τοποθετούνται στα άνω άκρα του ασθενούς. Τα σήματα τεχνηέντως παρακάμπτουν τις νευρικές οδούς προς το νωτιαίο μυελό και καταλήγουν απευθείας στα άκρα.⁸⁴

Μέσω συνεχών δοκιμών και έπειτα από πολλές λανθασμένες εκτιμήσεις, οι ερευνητές κατόρθωσαν να εντοπίσουν τον τρόπο με τον οποίο μπορούν να ρυθμίσουν την νευρική ώση ώστε να ανακτήσουν την κίνηση στα τραυματισμένα άκρα. Έπειτα από δύο χρόνια θεραπείας και εκατοντάδες συνεδρίες, ο Ian Burkhardt έχει κατορθώσει να πραγματοποιήσει αρκετές μηχανικές, απαραίτητες στην καθημερινότητα κινήσεις, με το δεξί του άνω άκρο, καθώς το σύστημα τον «μαθαίνει» καλύτερα και αυτός βελτιώνεται από τη χρήση του.



Εικόνα 4.12: Ο Ian Burkhardt ανακτά μερικώς την κίνησή του μέσω τεχνολογιών Big Data

Όπως αναφέρει ο Michael Schwemmer, επιστήμονας δεδομένων που δραστηριοποιείται στο πρόγραμμα, το σημαντικότερο στοιχείο είναι πως τόσο ο εγκέφαλος, όσο και οι μύες παραμένουν άθικτοι. Μόνο οι νευρικές οδοί παρουσιάζουν το πρόβλημα και εκεί πρέπει να εστιάσει η έρευνα. Το σημαντικότερο εργαλείο που έχουν στη διάθεσή τους και έχει συμβάλλει στα εκπληκτικά αυτά αποτελέσματα όπως, μάλιστα επισημαίνει ο ίδιος, είναι η αξιοποίηση των Big Data.. «Το πρόγραμμα αποτελεί μία άκρως ενδιαφέρουσα εφαρμογή της τεχνολογίας των Big Data Analytics, καθώς έχουμε τη δυνατότητα να διαχειριστούμε ροές Big Data που καταφθάνουν με μεγάλη ταχύτητα (Velocity).

Το βασικότερο μειονέκτημα του συστήματος είναι το μέγεθος, το οποίο προσπαθούμε να μειώσουμε».

Επιπλέον, μέσω των πειραμάτων, έχει επιβεβαιωθεί πως τα συναισθήματα του ασθενούς επηρεάζουν τα σήματα που στέλνει ο εγκέφαλός του, κάτι το οποίο επιτάσσει την ανάγκη καθημερινής βαθμονόμησης και προσαρμογής του συστήματος στη διάθεση του ασθενούς και ουσιαστικά καθιστά δύσκολη υπόθεση την οικιακή χρήση. Σε αυτό το εμπόδιο φιλοδοξεί να δώσει λύση η τεχνολογία των Big Data, προσαρμόζοντας αυτόματα το σύστημα στα συναισθήματα των ασθενών, διαχειριζόμενο σε πραγματικό χρόνο τους τεράστιους όγκους δεδομένων.

Το σύστημα αναμένεται στο μέλλον να εφαρμοσθεί και σε άτομα που πάσχουν από τη νόσο Parkinson.

4.6.2 Πρόγραμμα φροντίδας πρόωρα γεννημένων βρεφών με χρήση Big Data Analytics

Παρά το μικρό του μέγεθος, κάθε βρέφος παράγει μεγαλύτερο όγκο δεδομένων από αυτό που μπορεί να επεξεργαστεί και να «αποκρυπτογραφήσει» ένα νοσοκομειακό πληροφοριακό σύστημα. Στο πλαίσιο αυτό εγκαινιάστηκε το 2013 στον Καναδά το πρόγραμμα **Artemis**, μία ερευνητική σύμπραξη μεταξύ του Πανεπιστημίου University of Ontario Institute of Technology, της IBM, και της Παιδιατρικής Κλινικής του Τορόντο.⁸⁵ Έπειτα, έλαβε τη στήριξη πολλών οργανισμών μεταξύ των οποίων του Canada Foundation for Innovation και του Canadian Institutes for Health Research. Μάλιστα το εγχείρημα έλαβε χρηματοδότηση 3 εκατομμυρίων δολλαρίων ώστε να δημιουργηθεί μία εμπορική πλατφόρμα η οποία θα διατεθεί στην αγορά για την καλύτερη φροντίδα των πρόωρα γεννημένων βρεφών.

Σύμφωνα μάλιστα με τον Παγκόσμιο Οργανισμό Υγείας (**World Health Organization** ή **WHO**), οι πρόωρες γεννήσεις αποτελούν την κύρια αιτία θνησιμότητας βρεφών παγκοσμίως. Περίπου το 25% των πρόωρων βρεφών μολύνονται από λοιμώξεις και το 10% αυτών καταλήγουν από διάφορες επιπλοκές.

Στα πλαίσια της προσπάθειας αυτής παρακολουθούνται επί του παρόντος περισσότερα από 1.000 πρόωρα γεννημένα βρέφη. Το περιβάλλον του Artemis, που είναι χτισμένο σε πλατφόρμα της IBM,

συνδέεται με τα νοσοκομεία Women & Infants Hospital in Providence, R.I. και Children's Hospital of Fudan University in Shanghai στην Κίνα.

Για τα νοσοκομεία που φροντίζουν πρόωρα βρέφη, η συνεισφορά από το πρόγραμμα Artemis είναι μεγάλη. Το εγχείρημα αυτό έχει αποδείξει μέσω της ιατρικής έρευνας ότι:

- Μπορεί να συνδυάσει πληροφορίες από τους καρδιακούς παλμούς και την αναπνοή με άλλα φυσιολογικά δεδομένα για τη μείωση των ψευδών θετικών ενδείξεων (false positive) σε ελέγχους ανίχνευσης σηψαιμίας σε σύγκριση με την αποκλειστική χρήση πληροφοριών καρδιακών παλμών.
- Μπορεί να εντοπίσει αυτόματα το είδος της άπνοιας ενός βρέφους με ακρίβεια μεγαλύτερη από 98%.
- Μπορεί αυτόματα να διαχωρίζει τις καταστάσεις ύπνου και αφύπνισης στα νεογνά για να βοηθήσει τους γιατρούς να εκτιμήσουν τον τρόπο ανάπτυξης του εγκεφάλου ενός βρέφους.
- Μπορεί να παρέχει πληροφορίες σχετικά με τα επίπεδα οξυγόνου που είναι σε θέση να χρησιμοποιηθούν ώστε τα βρέφη να μην λαμβάνουν υπερβολικές ποσότητες οξυγόνου, γεγονός που μπορεί να οδηγήσει σε μόνιμη οφθαλμική βλάβη.

Σε νεογνικά τμήματα, το χάσμα ανάμεσα στα δεδομένα και τις πληροφορίες που μπορούν να αξιοποιηθούν δεν θα μπορούσε να είναι ευρύτερο, υποστηρίζει ο Carolyn McGregor, διδάκτωρ της πληροφορικής στην υγεία. Σημειώνει ότι «δεν υπάρχει έλλειψη δεδομένων για τα πρόωρα βρέφη, αφού οι καρδιές τους χτυπούν έως και 8.000 φορές και παίρνουν περίπου 2.000 αναπνοές την ώρα».⁸⁵

Υποστηρίζει πως ανακάλυψε το κενό που υπάρχει στην αξιοποίηση των δεδομένων όταν κλήθηκε να προτείνει τρόπους βελτίωσης του συστήματος: «Όλα αυτά τα δεδομένα που παράγονται κάθε δευτερόλεπτο ήταν απλά αριθμοί σε μια οθόνη παρακολούθησης. Οι άνθρωποι εξετάζουν τη φροντίδα που θεωρούν ότι είναι πολύ προηγμένη, αλλά αυτό που τους παρέχεται είναι ένας τεράστιος όγκος δεδομένων ο οποίος δεν είναι σε μορφή που μπορούν εύκολα να χρησιμοποιήσουν. Δεν υπήρχε αρχιτεκτονική σχεδίαση που θα μπορούσε να απορροφήσει όλα αυτά τα δεδομένα και να τα χρησιμοποιήσει αποτελεσματικά. Όλες αυτές οι πληροφορίες ανά δευτερόλεπτο μετρώνται με σκοπό

να δημιουργήσουν μία ένδειξη που υποτίθεται ότι υποδηλώνει την κατάσταση ενός ασθενούς, η οποία όμως χάνεται, εξαιτίας του πλούτου της». ⁸⁵ Μια ιατρός το νοσοκομείου παραδέχτηκε πως βασιζόταν συχνά στο ένστικτο των κατάλληλα εκπαιδευμένων νοσοκόμων, αντί των δεδομένων παρακολούθησης, για να εντοπίσει βρέφη που πιθανώς κινδύνευαν από κάποιο σοβαρό πρόβλημα υγείας.

Ο McGregor θέλησε να γεφυρώσει αυτό το χάσμα σχεδιάζοντας ένα σύστημα Big Data για να αναλύσει τεράστιες ποσότητες δεδομένων υψηλής ταχύτητας, να εντοπίσει αν οι αλλαγές στα δεδομένα υγείας σχετίζονται με ορισμένες συνθήκες και να προειδοποιήσει νοσηλευτές και ιατρούς για τις επικείμενες απειλές για την υγεία του βρέφους. Για παράδειγμα, το πρόγραμμα Artemis έχει εντοπίσει πρότυπα σε πραγματικό χρόνο για τις συνθήκες που μπορούν να αναπτύξουν τα βρέφη στη μονάδα εντατικής θεραπείας νεογνών, όπως:

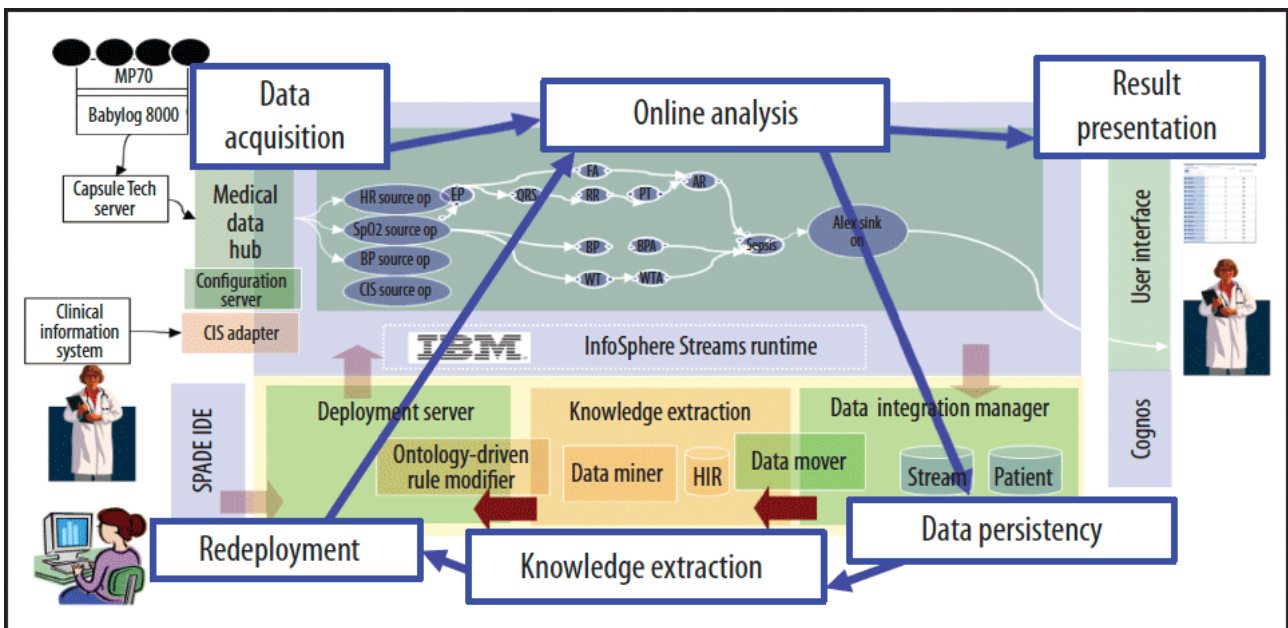
- Παύσεις στην αναπνοή (άπνοια) λόγω πρόωρου τοκετού.
- Η αμφιβληστροειδοπάθεια της πρόωρης γέννησης (βλάβη των οφθαλμών).
- Αναιμία του πρόωρου τοκετού (ανεπάρκεια ερυθρών αιμοσφαιρίων ως αποτέλεσμα της συχνής λήψης αίματος για εξετάσεις).

Το έργο χρησιμοποιεί τεχνικές εξόρυξης, κατοχυρωμένες με δίπλωμα ευρεσιτεχνίας από τον McGregor, οι οποίες έχουν σχεδιαστεί για την εξαγωγή μη τετριμμένων και δυνητικά χρήσιμων αφηρημένων πληροφοριών από μεγάλες συλλογές δεδομένων, στην περίπτωση των οποίων τα αριθμητικά δεδομένα παράγονται από συσκευές παρακολούθησης.

Το σύστημα ανάλυσης πραγματοποιεί αφαιρέσεις από τα δεδομένα εισόδου και αναζητά επαναλαμβανόμενα μοτίβα. Στη συνέχεια, ελέγχει εάν οι ασθενείς που είχαν προβλήματα υγείας, όπως μολύνσεις, σύνδρομο αναπνευστικής δυσχέρειας ή μορφές άπνοιας στον ύπνο, εμφάνιζαν τα ίδια μοτίβα δεδομένων στη φυσιολογική τους κατάσταση. Αφού τα μοντέλα αυτά προσδιοριστούν ως σημαντικά για την υγεία και αποδειχθούν οι σχέσεις με συγκεκριμένες συνθήκες, το σύστημα ανάλυσης μπορεί να αναζητήσει αυτά τα είδη μοτίβων σε ολόκληρο τον πληθυσμό των ασθενών και να προειδοποιήσει τους νοσηλευτές και τους ιατρούς όταν τα δεδομένα δείχνουν ότι ένα μωρό κινδυνεύει από κάποιο πρόβλημα υγείας.

Το πρόγραμμα Artemis χρησιμοποιεί τρία συστήματα ιατρικής συνδεσιμότητας από τα κλινικά κέντρα Capsule Tech, ExcelMedical και True Process, για την τροφοδοσία δεδομένων σε πραγματικό χρόνο σε μια βάση δεδομένων με βάση το υπολογιστικό νέφος (cloud computing) και ένα περιβάλλον ανάλυσης που βασίζεται στην πλατφόρμα InfoSphere και στη σχεσιακή βάση δεδομένων DB2, αμφότερα προϊόντα της IBM.

Για την ασφάλεια και αξιοπιστία των δεδομένων, υπάρχουν δύο αντίγραφα τους. Το πρώτο βρίσκεται στο νοσοκομείο και το άλλο αποστέλλεται μέσω συνδέσεων σε πραγματικό χρόνο σε περιβάλλον υπολογιστικού νέφους προς επεξεργασία. Οι πληροφορίες στην αποθήκη δεδομένων χρησιμοποιούνται για αναδρομική ανάλυση εφαρμόζοντας τεχνικές εξόρυξης δεδομένων (data mining) εντός συγκεκριμένου χρονικού διαστήματος ώστε να καταλήξουν σε υποθέσεις που μπορούν να αξιολογηθούν από τους ερευνητές. Αφού αποδειχθεί ότι οι υποθέσεις αυτές είναι βάσιμες και αληθείς, μπορούν να εφαρμοσθούν ως κανόνες στα συστήματα παρακολούθησης των νοσοκομείων.



Εικόνα 4.13: Λειτουργικό διάγραμμα εγχειρήματος Artemis

Η δυνατότητα του έργου Artemis να αναλύει τα φυσιολογικά δεδομένα των ασθενών με συχνότητα δευτερολέπτου σε πληροφορίες σχετικές με στοιχεία που αφορούν την υγεία τους, επεκτείνεται ήδη πέρα από το νοσοκομείο. Έχει ήδη δρομολογηθεί συνεργασία με διαστημικές υπηρεσίες για να

εφαρμοστούν οι τεχνικές του Artemis για την παρακολούθηση αστροναυτών και επίσης να μελετηθεί ο τρόπος με τον οποίο οι καταστάσεις άγχους μεταβάλλουν την ανθρώπινη συμπεριφορά.

Για τα διαστημικά προγράμματα, οι αστροναύτες στον Διεθνή Διαστημικό Σταθμό έχουν εβδομαδιαίες εξετάσεις. Ο McGregor συνεργάζεται με τους διαστημικούς οργανισμούς και ειδικότερα με τον ρωσικό διαστημικό οργανισμό, για να συνδυάσει τις τεχνικές παρακολούθησης με έναν ευέλικτο αλγόριθμο ανάλυσης Big Data, που ανέπτυξαν οι Ρώσοι για να παράσχουν μια πραγματική εικόνα της συνολικής υγείας των αστροναυτών και κοσμοναυτών. Μετρώνται μεγέθη, όπως οι μεταβολές της ποσότητας υγρών του σώματος, η καταπόνηση των οστών και άλλων προβλημάτων υγείας που σχετίζονται με ένα περιβάλλον χωρίς βαρύτητα. Η προσπάθεια έχει σχεδιαστεί για να καταδείξει πώς θα μπορούσε να εφαρμοστεί το εγχείρημα Artemis για την καλύτερη παρακολούθηση της υγείας των ανθρώπων στο Διεθνή Διαστημικό Σταθμό καθώς και η προτεινόμενη αποστολή της NASA για το 2030 στον Άρη. Τα αποτελέσματα της έρευνας αναμένεται να καρπωθούν και άλλες ερευνητικές προσπάθειες, στις οποίες οι διαστημικοί ταξιδιώτες προβλέπεται να στερηθούν την άμεση ιατρική φροντίδα για μεγάλα χρονικά διαστήματα.

Επιπλέον, ο McGregor συνεργάστηκε με στρατιωτικές ομάδες για να αναλύσει τα ιατρικά τους δεδομένα ενώ εκπαιδεύονται σε παιχνίδια εικονικής πραγματικότητας. Το έργο επιδιώκει να κατανοήσει ποιες αλλαγές στην καρδιακή συχνότητα και άλλες μεταβολές μπορούν να οδηγήσουν σε προβλήματα υγείας, όπως η λιποθυμία κατά τη διάρκεια των εγχειρήσεων, καθώς και ποια δεδομένα υγείας συνδέονται με το σύνδρομο μετατραυματικού στρες.

4.6.3 Καλύτερη πρόβλεψη εμφάνισης καρκίνου του μαστού

Η πρόβλεψη των αποτελεσμάτων της υγείας μέσω των Big Data Analytics συμβάλλει στους στόχους της υγειονομικής περίθαλψης. Μάλιστα, μια νέα μέθοδος βασισμένη σε Big Data έχει αποδειχθεί ότι βελτιώνει σημαντικά το ποσοστό πρόβλεψης του καρκίνου του μαστού. Η μέθοδος αναπτύχθηκε από μια ερευνητική ομάδα του Πανεπιστημίου Princeton, ονομάζεται βαθμολογία επίδρασης (**Influence Score** ή **I-Score**) και βελτιώνει το ποσοστό πρόβλεψης σε δεδομένα πραγματικών ασθενειών.⁸⁶

Τα αποτελέσματα της μελέτης, που δημοσιεύτηκαν το Δεκέμβριο του 2016 στο περιοδικό Proceedings of the National Academy of Sciences, έδειξαν ότι το I-score βελτίωσε το ποσοστό πρόβλεψης στα στοιχεία για τον καρκίνο του μαστού από 70% έως 92%. Ισοδύναμα, το I-score είναι

σε θέση να μειώσει το ποσοστό σφάλματος για την ορθή πρόβλεψη του καρκίνου του μαστού από 30% σε μόλις 8%.

Η λειτουργία του εγχειρήματος βασίζεται στην αξιοποίηση των Big Data, που προκύπτουν από διάφορες εξετάσεις και μετρήσεις. Συγκεκριμένα, επιτυγχάνεται ο προσδιορισμός αλληλεπιδράσεων υψηλότερης τάξης, που προκύπτουν με πολύπλοκο τρόπο μεταξύ πολλών μεταβλητών. Ουσιαστικά, εντοπίζεται μέσω της διαδικασίας αυτής ένα σύνολο μεταβλητών που κάνει εκτίμηση του κινδύνου με αποδοτικό και αξιόπιστο τρόπο. Η εφαρμογή της προσέγγισης στα πραγματικά δεδομένα για την ασθένεια δεν είναι μόνο επιτυχής στην εύρεση συνόλων μεταβλητών για την ασθένεια, αλλά έχει συμβάλει στην ανακάλυψη νέων προγνωστικών μεταβλητών που δεν είχαν εμφανιστεί ως παράγοντες χρήσιμοι για τη διάγνωση στις παραδοσιακές μεθόδους της Ιατρικής.

Σύμφωνα με την Adeline Lo, μεταδιδακτορική ερευνήτρια του Princeton, η μέθοδος είναι εξαιρετικά αποδοτική για μεγάλα σύνολα δεδομένων, όπως τα γενετικά, το οποίο σημαίνει ότι μπορεί να βρει εφαρμογή και σε πολλούς ακόμα τομείς της υγείας. Η προσέγγιση που υιοθετεί, το καθιστά ισχυρό εργαλείο για την ανάλυση πολυδιάστατων δεδομένων, ετερογενών πηγών, ενώ η χρησιμότητά του εκτείνεται και πέραν των ορίων του τομέα της υγειονομικής περίθαλψης.

4.6.4 Πρόβλεψη εξάπλωσης της επιδημίας του ιού Zika

Ο μεταδιδόμενος μέσω των κουνουπιών ιός Zika πρωτοεμφανίσθηκε στην Ουγκάντα της Αφρικής το 1947 και το 2007 γνώρισε μία μετάλλαξη, η οποία στα τέλη του 2015, εντοπίστηκε στην ήπειρο της Αμερικής. Μέσα σε πέντε μόλις μήνες, η Βραζιλία επιβεβαίωσε 3.000 περιπτώσεις εγκύων γυναικών που μολύνθηκαν από τον ιό, ενώ η ασθένεια εξαπλώθηκε και σε ορισμένες Πολιτείες των ΗΠΑ.⁸⁷

Ο Kamran Khan, ιατρός και επιστήμονας με ειδίκευση στις επιδημίες στο St. Michael Hospital υποστηρίζει ότι το μόνο βέβαιο με την εξάπλωση μολυσματικών ασθενειών είναι: «Αν αρχίσετε να αναλύετε την κατάσταση όταν εμφανιστεί ένα ξέσπασμα, είναι ήδη πολύ αργά». Ο Khan έχει περάσει τη σταδιοδρομία του καταπολεμώντας τους ιούς Zika, Ebola, Lassa Fever και άλλες θανατηφόρες και μακροχρόνιες μολυσματικές ασθένειες που έχουν αναπτυχθεί σε απροσδόκητα σημεία και έχουν προκαλέσει δημόσιο πανικό και θάνατο.

Αφού, εντόπισε το κενό που υπάρχει στην αξιοποίηση των δεδομένων, ο Khan έγινε ο ιδρυτής της **BlueDot**, μιας κοινωνικής επιχείρησης με έδρα το Τορόντο, που επικεντρώνεται στην αξιοποίηση των Big Data Analytics για την αποτελεσματικότερη πρόβλεψη μολυσματικών ασθενειών.

Η πλατφόρμα **BioDiaspora** που τροφοδοτεί τις προσπάθειες της BlueDot αναπτύχθηκε το 2008 στο Ινστιτούτο Li Ka Shing Knowledge Institute του St. Michael's Hospital. Το 2013, η πλατφόρμα μετατράπηκε σε εταιρεία κερδοσκοπικού χαρακτήρα και το 2014 μετονομάστηκε σε BlueDot. Η εταιρεία, η οποία έλαβε χρηματοδότηση από το ινστιτούτο Li Ka Shing και τεχνολογικούς επενδυτές, συνδυάζει τις μελέτες ειδικών επιστημόνων για τις μολυσματικές ασθένειες, επιστημόνων δεδομένων, ερευνητών και μηχανικών πληροφορικής με δεδομένα πραγματικού χρόνου από δρομολόγια εμπορικών πτήσεων, τα οποία μπορεί να σχετίζονται με τον πληθυσμό ανθρώπων, ζώων και εντόμων, πληροφορίες σχετικά με το κλίμα μέσω των δορυφόρων, καθώς και εκθέσεις για την εμφάνιση ασθενειών.

Η BlueDot έχει συνάψει συμφωνία συνεργασίας πενταετούς διάρκειας με το Centers for Disease Control and Prevention, ενώ έχει λάβει και χρηματοδότηση από το Υπουργείο Εξωτερικών, Εμπορίου και Ανάπτυξης του Καναδά. Βασικός στόχος είναι να υποστηρίξει τους στόχους της Παγκόσμιου Οργανισμού Υγείας στην Δυτική Αφρική και να δημιουργήσει προοπτικές ελέγχου μολυσματικών ασθενειών. Η BlueDot αναπτύσσει ερευνητικές εκθέσεις και μοντέλα κινδύνου που διανέμονται σε φορείς δημόσιας υγείας και άλλους παρόχους υπηρεσιών υγειονομικής περίθαλψης για να περιγράψουν τη δυνητική εξάπλωση επιδημιών και τις εκτιμώμενες επιπτώσεις στους πληθυσμούς που βρίσκονται σε κίνδυνο. Επίσης, αναπτύσσει εργαλεία και εφαρμογές διαδικτύου, τα οποία μπορούν να βοηθήσουν τους πρώτους παρατηρητές νέων ασθενειών να τις αναφέρουν και να βοηθήσουν τους οργανισμούς δημόσιας υγείας να αναπτύξουν ταχύτερα μηχανισμούς αντιμετώπισης.

Η BlueDot και οι συνεργάτες της από τα Πανεπιστήμια Harvard και Oxford, έλαβαν προειδοποίηση σχετικά με την εξάπλωση του ιού Zika στη Βραζιλία και άμεσα εξέδωσαν ένα έντυπο που δημοσιεύθηκε στο ιατρικό περιοδικό Lancet τον Ιανουάριο του 2016, ένα μήνα προτού ο Παγκόσμιος Οργανισμός Υγείας αναγνωρίσει την επικινδυνότητα του ιού. Ως έκτακτης ανάγκης μέτρα για την δημόσια υγεία τον Φεβρουάριο του ίδιου έτους, τα Κέντρα Ελέγχου και Πρόληψης Νοσημάτων εξέδωσαν προειδοποιήσεις για το κοινό.

Η αφορμή του εγχειρήματος του Khan ήταν το ξέσπασμα του σοβαρού οξέος αναπνευστικού συνδρόμου (SARS) το 2003 στην πατρίδα του, το Τορόντο. Η ασθένεια πρωτοεμφανίστηκε στην επαρχία Γκουανγκντόνγκ στην Κίνα το Νοέμβριο του 2002. Ένας ιατρός που παρακολουθούσε ασθενείς του ιού SARS, ταξίδεψε στο Χονγκ Κονγκ και ανέπτυξε τα πρώτα συμπτώματα. Κατά τη διάρκεια της παραμονής του, μολύνει 12 άλλους επισκέπτες του ξενοδοχείου, συμπεριλαμβανομένης μιας 78χρονης γυναίκας από τον Καναδά που αργότερα επέστρεψε στο Τορόντο. Η μετάδοση του SARS οδήγησε σε ένα ξέσπασμα που ξεκίνησε τον Μάρτιο του 2003, με αποτέλεσμα 44 άτομα στον Καναδά να πεθάνουν από τον ιό, περίπου 400 να νοσήσουν και 25.000 κάτοικοι του Τορόντο να βρίσκονται σε καραντίνα. Τα στοιχεία αυτά είναι ενδεικτικά για την ταχύτητα εξάπλωσης των επιδημιών και επιβεβαιώνουν την ανάγκη επιστράτευσης κάθε δυνατού μέσου, ώστε να περιοριστούν έγκαιρα.

Η μεγαλύτερη πρόκληση για την BlueDot δεν είναι ο όγκος και η ταχύτητα των δεδομένων, αλλά η ποικιλία τους. Τα σύνολα δεδομένων που χρησιμοποιούνται από δορυφορικές τροφοδοσίες σε πραγματικό χρόνο για κλιματολογικά δεδομένα, οι κλινικές αναφορές και τα δρομολόγια ταξιδιών, δεν είχαν συνδυαστεί ποτέ στο παρελθόν. Το μεγαλύτερο εμπόδιο ήταν να οργανωθούν και να επεξεργαστούν, ώστε να είναι διαθέσιμα για διαφορετικές αναλύσεις, όταν παρουσιαστεί η ανάγκη. Αφού αρθεί αυτό το εμπόδιο, οι ερευνητές μπορούν να επικεντρωθούν σε πρόσθετα σύνολα δεδομένων ειδικά για μολυσματικές ασθένειες. Για παράδειγμα, έχουν τη δυνατότητα να αναλύσουν τον ιό Zika συμπεριλαμβάνοντας πληροφορίες για τον κύκλο ζωής του κώνωπα που μεταφέρει τον ιό, στοιχεία σχετικά με το κλίμα που επικρατούσε την περίοδο της εξάπλωσης και με ποιο τρόπο την επηρέασε, καθώς και δεδομένα για την ανθρώπινη κινητικότητα και τις περιοχές που ενδέχεται να τεθούν άμεσα υπό κίνδυνο.

Όλα τα εισερχόμενα δεδομένα στην πλατφόρμα είναι χρονικά και γεωγραφικά ομαδοποιημένα, και τροφοδοτούν ένα διαδικτυακό σύστημα γεωγραφικών πληροφοριών από το ερευνητικό Ινστιτούτο ESRI (Environmental Systems Research Institute). Οι ερευνητές στην BlueDot μπορούν στη συνέχεια να δημιουργήσουν απεικονίσεις της εξάπλωσης της νόσου, καθώς και να χρησιμοποιήσουν αυτά τα δεδομένα με σκοπό την εκτέλεση προγνωστικών μοντέλων για να κατανοήσουν πότε και πού θα εμφανιστούν περαιτέρω εστίες του ιού, σε ποιες περιόδους, σε πόσα άτομα, και πολλές άλλες μεταβλητές που θεωρούνται χρήσιμες σε κάθε περίπτωση.

Όσον αφορά τον ιό Zika, ο χάρτης κινδύνου της BlueDot μέσω του εντύπου που δημοσιεύθηκε στην Lancet, προέβλεπε την εξάπλωση της νόσου στη Βραζιλία και επίσης προειδοποίησε ότι η ασθένεια

πιθανότατα θα εξαπλωθεί στις Ηνωμένες Πολιτείες μέσω της Φλόριντα, με βάση τον όγκο των ταξιδιών με αυτό τον προορισμό, το κλίμα και τα είδη των κουνουπιών που μεταφέρουν τον ιό. Παράλληλα, το έγγραφο επισήμαινε ότι η καλύτερη στέγαση και το λιγότερο στάσιμο νερό στη Φλόριντα, σε σύγκριση με τις πληγείσες περιοχές της Βραζιλίας, θα είχαν ως αποτέλεσμα την περιορισμένη μετάδοση στην Πολιτεία αυτή.

Το γεγονός ότι όλες οι εκτιμήσεις της BlueDot αποδείχθηκαν αληθείς, επιβεβαιώνει την αξία των Big Data στον περιορισμό των επιδημιών, αποτελώντας ένα ακόμα παράδειγμα όπου η χρήση τους συμβάλλει στην ανθρώπινη ύπαρξη και υγεία. Δημιουργούνται νέες προοπτικές στον τομέα της υγείας και το μόνο βέβαιο είναι ότι το πεδίο δράσης θα εξαπλωθεί σε παγκόσμιο επίπεδο για τον περιορισμό ασθενειών και επιδημιών.

ΚΕΦΑΛΑΙΟ 5

ΤΕΧΝΙΚΕΣ ΚΑΙ ΟΡΓΑΝΩΤΙΚΕΣ ΠΡΟΚΛΗΣΕΙΣ

5.1 Αποθετήρια Δεδομένων (Data Silos)

Αν και έχει ήδη αναφερθεί πως τα διαθέσιμα δεδομένα για την υγεία αναπτύσσονται με εκθετικό ρυθμό, η πλειοψηφία τους βρίσκεται σε επιμέρους αποθετήρια, ένα φαινόμενο για το οποίο έχει επικρατήσει ο όρος **Data Silos**.⁸⁸ Στην ουσία, πρόκειται για αποθετήρια δεδομένων, τα οποία παραμένουν στη διάθεση ενός οργανισμού ή ακόμα και σε επιμέρους τμήματα οργανισμών και δεν διατίθενται εξωτερικά. Τα data silos είθισται να παρουσιάζονται ως φυσικό επακόλουθο της λειτουργίας μεγάλων οργανισμών, όπου υπάρχει έντονος ανταγωνισμός μεταξύ των τμημάτων. Είναι γενικά αποδεκτό πως αποτελούν σημαντικό εμπόδιο για την αξιοποίηση των Big Data και υπάρχει μεγάλη προσπάθεια αντιμετώπισης του φαινομένου και καλλιέργειας των εννοιών της συνεργασίας και του κοινού σκοπού.

Συγκεκριμένα, στον κλάδο της υγείας, τα δεδομένα που συλλέγονται από μια κλινική ή από ένα νοσοκομείο ως επί το πλείστον διατηρούνται εντός των ορίων του παρόχου υγειονομικής περίθαλψης. Επιπλέον, τα δεδομένα που αποθηκεύονται σε ένα νοσοκομειακό πληροφοριακό σύστημα, δεν είναι σχεδόν ποτέ εντασσόμενα σε ένα ενιαίο σύστημα. Για παράδειγμα, αν λάβουμε υπόψη όλα τα διαθέσιμα δεδομένα σε ένα νοσοκομείο από την προοπτική ενός ασθενούς, πληροφορίες σχετικά με τον ασθενή θα υπάρχουν στις βάσεις δεδομένων του Ηλεκτρονικού Φακέλου Υγείας (Electronic Medical Record), στα εργαστήρια καθώς και σε συστήματα απεικόνισης και ιατρικές εικόνες. Πληροφορίες που περιγράφουν το ποιοι ιατροί και νοσηλευτές ήρθαν σε επαφή με το συγκεκριμένο ασθενή, θα υπάρχουν επίσης κατοχυρωμένα. Ωστόσο, στη συντριπτική πλειονότητα των περιπτώσεων, τα αποθετήρια των δύο αυτών δεδομένων βρίσκονται σε διαφορετικές βάσεις δεδομένων. Έτσι προκύπτουν δυσκολίες αξιοποίησης των δεδομένων αφού δεν βρίσκονται σε ενιαία δομή στο στάδιο αυτό.

Είναι επίσης σημαντικό να καταστεί σαφές ότι στο σύγχρονο κόσμο τα ιατρικά δεδομένα του ασθενούς δεν παραμένουν μόνο εντός των ορίων ενός παρόχου υγειονομικής περίθαλψης. Ο πάροχος ιατρικής ασφάλισης και οι φαρμακευτικές βιομηχανίες κατέχουν επίσης πληροφορίες σχετικά με τις παθήσεις των ασθενών και τα χαρακτηριστικά των συνταγογραφούμενων φαρμάκων, αντίστοιχα. Επίσης, όλο και συχνότερα, δεδομένα ασθενών που δημιουργούνται από συσκευές του Διαδικτύου των Πραγμάτων (Internet of Things), όπως μετρητές φυσικής κατάστασης και γυμναστικών επιδόσεων, συσκευές παρακολούθησης της πίεσης του αίματος και πολλές ακόμα, παρέχουν σημαντικές πληροφορίες σχετικά με τα χαρακτηριστικά του καθημερινού τρόπου ζωής ενός ατόμου.

Διαφορετικά ευρήματα προέρχονται από τη σύνδεση μεταξύ των δεδομένων του Ηλεκτρονικού Ιατρικού Αρχείου, των εργαστηριακών δεδομένων, των πληροφοριών φαρμακευτικών αγωγών, των συμπτωμάτων και την ομαδοποίησή τους από ότι τα επιμέρους δεδομένα ξεχωριστά. Αν σε αυτά προστεθούν ιατρικές σημειώσεις, εξιτήρια και ημερολόγια ασθενών, ιατρικές έρευνες και γενικότερα σύνδεση δομημένων με μη δομημένα δεδομένα μπορούν να προκύψουν πολλαπλά οφέλη, όπως μείωση χρόνιων ασθενειών, εξατομικευμένα μοντέλα πρόληψης και μείωση του κόστους των υγειονομικών υπηρεσιών.

Οι μεταβάσεις για την υγειονομική περίθαλψη από ανεξάρτητες συλλογές όγκων δεδομένων σε ενιαία μοντέλα πέραν του χρόνου και της προσπάθειας, απαιτεί και τη διάθεση πόρων. Είναι, λοιπόν, σημαντικό για τους φορείς να έχουν πλήρη και ορθή αντίληψη των ιατρικών στοιχείων συγκεκριμένων πληθυσμιακών ομάδων. Ο μόνος τρόπος για να επιτευχθεί αυτό είναι να έχουν τη δυνατότητα συγκέντρωσης ανάμοιων συνόλων δεδομένων, στο πλαίσιο της εσωτερικής τους υποδομής, όπως μίας κλινικής ή ενός νοσοκομείου, από διαφορετικές πηγές. Αν ικανοποιηθεί η απαίτηση αυτή, θα υπάρξει μίας σχέση αμφίδρομου οφέλους, τόσο για τις μονάδες υγείας, όσο και για τους ασθενείς, θέτοντας το τρίπτυχο ποιότητα-κόστος-πρόσβαση σε νέα βάση.

5.2 Πολιτικές, κοινωνικές και βιοηθικές επιπτώσεις

Η εισαγωγή των Big Data, σε συνδυασμό με τη χρήση των μέσων κοινωνικής δικτύωσης και του διαδικτύου προκαλούν αλλαγές σε τομείς όπως η κοινωνία, η οικονομία και η διοίκηση, βελτιώνοντας τον τρόπο λήψης αποφάσεων και τη διαδικασία επίλυσης προβλημάτων.¹⁷

Συνεπώς, οι επιπτώσεις της χρήσης των δεδομένων στις πολιτικές διοίκησης είναι σημαντικές και απαιτούν ένα διαρκώς προσαρμοζόμενο στις εξελίξεις θεσμικό πλαίσιο. Η χρήση των Big Data είναι δυνατόν να δημιουργήσει βλάβη στην ιδιωτικότητα και ενδεχομένως στα προσωπικά δεδομένα υγείας.⁸⁹ Η χρήση των μέσων κοινωνικής δικτύωσης στα Big Data Analytics, για την παρακολούθηση των επιδημικών εξάρσεων και την άμεση προειδοποίηση για τη λήψη μέτρων, μπορεί επίσης να προκαλέσει βλάβη στην ιδιωτικότητα.^{90 91}

Στα πλαίσια της παγκόσμιας βιοοικονομίας έχουν αναδειχθεί πρακτικές νέων τρόπων παραγωγής γνώσης, όπως η συλλογική καινοτομία με ευρείας κλίμακας κοινοπραξίες (**consortia**), οι οποίες συγκροτούν τους κλάδους «**Big Science**» και «**Big Ethics**». Αυτές οι πρακτικές ενδυναμώνουν τις προοδευτικές κοινωνικές αλλαγές και εμπλέκουν νομικές και βιοηθικές διαστάσεις σχετικά με τα ανθρώπινα δικαιώματα και την προστασία των προσωπικών δεδομένων.

Η αναλυτική των Big Data επιφέρει τη δυναμική μετασχηματισμού του τρόπου παραγωγής και διανομής της υγειονομικής φροντίδας με τη χρήση τεχνολογίας, ώστε να είναι δυνατή η εκμετάλλευση κλινικών και λοιπών δεδομένων με σκοπό να προσφέρει το έδαφος για αποφάσεις που βασίζονται στην πληροφορία και την τεκμηρίωση.

Η αυξανόμενη τάση για χρήση και εμπορική εκμετάλλευση των δεδομένων της υγείας εγείρει ευρεία συζήτηση, σε παγκόσμια κλίμακα, για τις επιπτώσεις της μαζικής παραγωγής προϊόντων γνώσης ως συνδυασμού επιστημονικών και ηθικών κοινοπραξιών.⁹²

5.2.1 Προστασία της Ιδιωτικότητας

Έχει ήδη γίνει σαφές πως η χρήση των Big Data παράγει υπερπολύτιμες πληροφορίες στον τομέα της υγείας. Από την παρούσα μελέτη δεν θα μπορούσε να παραληφθεί το ιδιαίτερα λεπτό ζήτημα της προστασίας των προσωπικών δεδομένων. Τα συστήματα υγείας διαθέτουν ευαίσθητα προσωπικά δεδομένα, η δημοσιοποίηση των οποίων μπορεί να έχει καταστροφικές συνέπειες για τους ασθενείς.

Στο παρελθόν, τα άτομα που είχαν προσβληθεί από τον ιό της ανθρώπινης ανοσοανεπάρκειας (Human Immunodeficiency Virus ή HIV) αντιμετώπιζαν σοβαρά ζητήματα κοινωνικής αποδοχής λόγω της ασθένειας. Στις Η.Π.Α., ο οργανισμός **HIPAA Privacy Rule** (Health Insurance Portability and Accountability Act) είναι υπεύθυνος για τη σύνταξη και τήρηση των νομικών προδιαγραφών σχετικά με τη χρήση των προσωπικών δεδομένων.⁹³

Αποτελεί τεράστια πρόκληση για την αξιοποίηση των Big Data, η συγκέντρωση τεράστιων συνόλων δεδομένων, πολλών εκ των οποίων προστατεύονται βάσει νόμου, με την παράλληλη τήρηση των νόμιμων διαδικασιών. Ορισμένα στοιχεία μπορούν άμεσα να προσδιορίσουν το άτομο. Για παράδειγμα, σε όλα τα συστήματα υγείας, κάθε ασθενής διαθέτει ένα προσωπικό αριθμό κοινωνικής ασφάλισης και πληροφορίες σχετικά με την ασφάλιση υγείας που διαθέτει. Διαθέτει επίσης και άλλα προσωπικά στοιχεία, όπως τόπο γέννησης, διεύθυνση διαμονής, ηλικία, φύλο, πληροφορίες σχετικά με διαγνώσεις ασθενειών και θεραπειών, στοιχεία τα οποία ουσιαστικά μπορούν έμμεσα να τον ταυτοποιήσουν. Γενικότερα, όσο περισσότερα στοιχεία υπάρχουν για κάποιο άτομο, τόσο ευκολότερος είναι ο προσδιορισμός του. Η προστασία των δεδομένων προϋποθέτει λοιπόν το φιλτράρισμα όχι μόνο των άμεσων, αλλά και των έμμεσων αναγνωριστικών. Αναπόφευκτα, αυτό έρχεται σε σύγκρουση με την ανάγκη πρόσβασης στα δεδομένα, καθώς έχουν τεράστια σημασία για τους σκοπούς της ανάλυσης.

Στον τομέα της υγείας δεν είναι λοιπόν εφικτό να ακολουθηθεί η πολιτική **Safe Harbor** που τηρείται σε άλλες περιπτώσεις και προβλέπει αφαίρεση 18 συγκεκριμένων πεδίων.⁹⁴ Η υπεύθυνη κοινή χρήση των δεδομένων απαιτεί μία πιο ριζοσπαστική προσέγγιση, η οποία στηρίζεται σε εκτιμήσεις ρίσκου. Χαρακτηριστικό παράδειγμα αποτελεί η μέθοδος **Expert Determination**, η οποία είναι βασισμένη σε στατιστικές αρχές ώστε να διατηρείται η ιδιωτικότητα της πληροφορίας.⁹⁵ Με επιστημονικά αποδεκτό τρόπο, γίνεται εκτίμηση της πιθανότητας προσδιορισμού του ατόμου μέσω των έμμεσων αναγνωριστικών, υπό την προϋπόθεση πως τα άμεσα παραμένουν απόρρητα. Στη συνέχεια μέσω διαφόρων τεχνικών, όπως η γενίκευση, η συσσώρευση και η τυχαιότητα, προσδιορίζεται το σύνολο εκείνο που συγκεντρώνει τα περισσότερα από τα διαθέσιμα πεδία δεδομένων, δίχως να παραβιάζεται το ανώτατο όριο πιθανότητας ταυτοποίησης, όπως αυτό διαμορφώνεται από διεθνώς αναγνωρισμένους φορείς όπως το Κέντρο Ελέγχου Ασθενειών (**Center for Disease Control** ή **CDC**).⁹⁶ Η ανάγκη αξιόπιστης μετατροπής των δεδομένων σε ανώνυμα, διατηρώντας την υψηλή ποιότητα, είναι αυτή που ώθησε πολλούς οργανισμούς (Institute of Medicine, HITRUST, PhUSE και

Canadian Council of Academies) να εισηγηθούν τη χρήση προσεγγίσεων βασισμένων σε εκτιμήσεις ρίσκου για την αξιοποίηση των Big Data.

Η λήψη αποφάσεων σχετικά με την τήρηση της ιδιωτικότητας αποτελεί ένα από τα μεγαλύτερα εμπόδια για οργανισμούς που εξετάζουν το ενδεχόμενο αξιοποίησης της τεχνολογίας των Big Data.

5.2.2 Προτεινόμενες πρακτικές

Ορισμένες έρευνες επιβεβαιώνουν πως όλο και περισσότερα άτομα είναι δεκτικά και πρόθυμα να μοιραστούν τα προσωπικά ιατρικά τους δεδομένα για σκοπούς υπηρεσίας της υγείας, αλλά αυτό δεν φαίνεται ικανό να αναιρέσει τα ζητήματα που εγείρονται.⁹⁷ Οι εταιρίες που αποσκοπούν στην αξιοποίηση των Big Data, οφείλουν να υιοθετήσουν αυστηρά καθορισμένες πολιτικές προστασίας και ασφάλειας των δεδομένων, καθώς δεν είναι μόνο οι ασθενείς που εκτίθενται από πιθανές παραλείψεις και διαρροές, αλλά και οι ίδιες οι εταιρίες ως υπεύθυνες για την κυριότητα των προσωπικών δεδομένων. Στο πλαίσιο αυτό καταγράφονται ορισμένες προτεινόμενες στρατηγικές για την ελαχιστοποίηση του ρίσκου:⁹⁸

Δημιουργία ομάδας προστασίας ιδιωτικότητας: Πρόκειται για μία ομάδα αποτελούμενη από άτομα διαφόρων τμημάτων εντός του οργανισμού. Κρίνεται απαραίτητη, κατ' ελάχιστο, η ύπαρξη ενός επιβλέποντα υπευθύνου, μίας ομάδας τεχνολογίας της πληροφορίας και νομικών εκπροσώπων. Ο ρόλος της ομάδας είναι η πρόβλεψη εσωτερικών και εξωτερικών κινδύνων και η διαμόρφωση κατάλληλων στρατηγικών.

Προσδιορισμός δομών εκτίμησης του ρίσκου: Η λήψη αποφάσεων σχετικά με το ποιος θα έχει πρόσβαση, σε ποια σύνολα δεδομένων, πώς μπορεί να τα διαχειρίζεται και να τα επεξεργάζεται, σηματοδοτούν, όπως σε όλα τα πληροφοριακά συστήματα, έτσι και στον τομέα της υγείας, την ανάγκη λήψης ορισμένων στρατηγικών αποφάσεων. Προκειμένου να γίνει εκτίμηση του ρίσκου, πρέπει να εξετάζονται τόσο τα δεδομένα, όσο και οι διαχειριστές τους. Η ύπαρξη τέτοιων δομών και η συνεχής αναθεώρησή τους, συνιστούν θεμέλιους λίθους για την προστασία της ιδιωτικότητας των ασθενών.

Παροχή εκπαίδευσης στο προσωπικό: Οι εργαζόμενοι που είναι υπεύθυνοι για την έγκριση αιτημάτων πρόσβασης σε δεδομένα, οφείλουν να έχουν την κατάλληλη εκπαίδευση σχετικά με τις

προδιαγραφές που οφείλουν να τηρούν και τον τρόπο με τον οποίο πρέπει να δρουν εναρμονισμένοι με τα νομικά πρότυπα. Προφανώς, η εκπαίδευση οφείλει να γίνεται ανά τακτά χρονικά διαστήματα.

Ανάπτυξη ευέλικτου λογισμικού με δυνατότητες κλιμάκωσης: Το πλήθος των δεδομένων και των πεδίων καθιστούν αδύνατη τη διαχείρισή τους με στατικά εργαλεία χωρίς δυνατότητες προσαρμοστικότητας. Για το λόγο αυτό επιβάλλεται η δυνατότητα αλλαγών στο λογισμικό, εφόσον απαιτείται. Ιδανικά, το ίδιο το λογισμικό μπορεί να παρέχει εργαλεία εκτίμησης του ρίσκου και ελέγχου της ασφάλειας, εξοικονομώντας χρόνο από άλλες παραγωγικές μονάδες.

5.3 Το ευρύτερο πεδίο των προκλήσεων

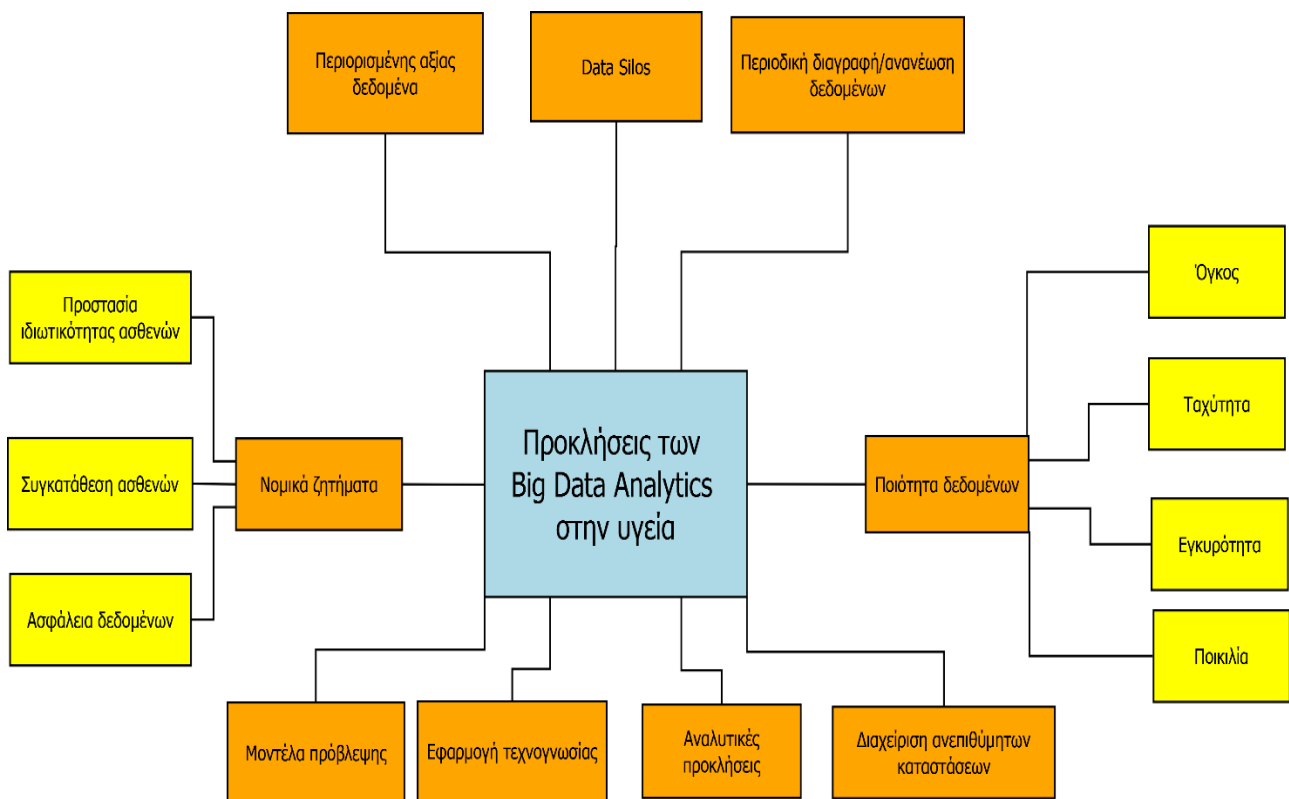
Οι προκλήσεις που προκύπτουν στα πλαίσια της αξιοποίησης της τεχνολογίας των Big Data Analytics είναι πολλές και ιδιαίτερα σημαντικές ώστε να καταστεί αποτελεσματική η προσπάθεια. Οι προκλήσεις είναι ετερογενείς και ποικιλόμορφες. Τα κυριότερα σημεία που πρέπει σε κάθε περίπτωση ένας φορέας που δραστηριοποιείται στο χώρο της υγείας να εξετάσει, στα πλαίσια αυτά, είναι:

- **Περιορισμένης αξίας δεδομένα:** Δεδομένα τα οποία ακόμα και αν αξιοποιηθούν κατάλληλα, δεν έχουν τη δυνατότητα να προσφέρουν κάποιου είδους αξία στον οργανισμό υγείας που χρησιμοποιεί τεχνολογίες Big Data Analytics. Για το λόγο αυτό είναι απαραίτητο να εξετάζονται ενδελεχώς τα αξιοποιήσιμα χαρακτηριστικά των διαθέσιμων δεδομένων, ούτως ώστε να υπάρχουν πραγματικές προοπτικές και συγκεκριμένη στόχευση.
- **Data Silos:** Η έλλειψη κοινού πνεύματος συνεργασίας μεταξύ οργανισμών, αλλά και εσωτερικά μεταξύ των επιμέρους τμημάτων, αναπόφευκτα παρεμποδίζει την κοινή χρήση δεδομένων. Πρέπει λοιπόν ο εκάστοτε φορέας να εξασφαλίζει την αποφυγή αυτού του κινδύνου, αναπτύσσοντας το κατάλληλο πνεύμα μεταξύ των εργαζομένων, γεγονός που δεν αποτελεί συνήθως μία τυπική διαδικασία.
- **Περιοδική διαγραφή/ανανέωση δεδομένων:** Είναι ένα καθαρά τεχνικό ζήτημα, το οποίο όμως μπορεί να δημιουργήσει δυσκολίες αν δεν τηρηθεί. Ουσιαστικά, αφορά τη διαχείριση των δεδομένων. Σε κάποιες περιπτώσεις υπάρχει ανάγκη περιοδικής διαγραφής των δεδομένων ή ανανέωσης τους, ενώ τα συστήματα που διατίθενται έχουν συγκεκριμένες δυνατότητες. Πρέπει λοιπόν να υπάρχει διαβεβαίωση ότι μπορεί να πραγματοποιείται δυναμική διαχείρισή των δεδομένων.
- **Ποιότητα δεδομένων:** Η ποιότητα των δεδομένων αφορά όλα τα βασικά χαρακτηριστικά που περιγράφουν τα Big Data, όπως έχουν παρουσιαστεί εκτενώς στο Κεφάλαιο 2. Συγκεκριμένα, διακρίνονται 4 κατηγορίες (4 V's) και συναντώνται οι ακόλουθες προκλήσεις:
 - **Όγκος:** Για την αποτελεσματική εκμετάλλευση των δεδομένων, οφείλει να εξασφαλιστεί η δυνατότητα διαχείρισης και αποθήκευσης του όγκου των δεδομένων, καθώς και να

προσδιορισθεί το μέγεθός τους. Απαιτούνται σχεδόν πάντα δυνατότητες κλιμάκωσης, καθώς οι ανάγκες αυξάνονται διαρκώς και κατ' επέκταση και ο απαιτούμενος προς αξιοποίηση όγκος.

- **Ταχύτητα:** Κάθε οργανισμός οφείλει να εξετάσει την ταχύτητα με την οποία μπορεί να αποθηκεύσει, επεξεργαστεί και αξιοποιήσει τα διαθέσιμα δεδομένα και να βελτιώνει συνεχώς τις επιδόσεις, κυρίως σε περιπτώσεις που η ταχύτητα προσέλευσης των δεδομένων είναι μεγάλη.
- **Εγκυρότητα:** Η διασφάλιση της εγκυρότητας των δεδομένων είναι απαραίτητη για τις ανάγκες του εγχειρήματος και αποτελεί απαιτητική διαδικασία.
- **Ποικιλία:** Ο προσδιορισμός όλων των πηγών δεδομένων, οι τεχνικές προκλήσεις που επιτάσσει η κάθε πηγή και η αποδοτική διαχείρισή τους, είναι αναπόσπαστο κομμάτι και μείζουσα πρόκληση για κάθε προσπάθεια ανάλυσης Big Data.
- **Διαχείριση ανεπιθύμητων καταστάσεων:** Επειδή υπάρχουν πολλές εμπλεκόμενες διαδικασίες, πέραν την πρόβλεψης και οργάνωσης των λειτουργιών, απαιτείται και μελέτη των απρόβλεπτων καταστάσεων, όπως τεχνικά σφάλματα που ενδέχεται να θέσουν το σύστημα εκτός λειτουργίας κ.ό.κ. Σε κάθε περίπτωση απαιτείται ένα σχέδιο δράσης και εκτίμησης κινδύνου για τις περιπτώσεις αυτές.
- **Αναλυτικές προκλήσεις:** Πέραν των τεχνικών απαιτήσεων, έχει τεράστια σημασία η «αποκρυπτογράφηση» των δεδομένων, η κατανόησή τους, η ανάπτυξη αναλυτικής σκέψης και μεθόδων για την δημιουργία αξίας. Συμβαίνει συχνά τα δεδομένα να ερμηνεύονται λανθασμένα από τον ανθρώπινο παράγοντα, γεγονός που ενδέχεται να επιφέρει αντίθετα αποτελέσματα από τα επιθυμητά.
- **Εφαρμογή τεχνογνωσίας:** Οι ανάγκες αυξάνονται διαρκώς στον τομέα της υγείας, μέσω νέων ερευνών, παρατηρήσεων, επιστημονικών άρθρων κ.ό.κ. Παράλληλα όμως αυξάνονται και οι τεχνολογικές δυνατότητες που είναι δυνατόν να συμβάλλουν στην κάλυψη των αναγκών. Για αυτό το λόγο, είναι απαραίτητη η γνώση των τεχνολογικών εξελίξεων και η παρέμβαση, όπου κρίνεται απαραίτητο, για την αντιμετώπιση εγγενών δυσκολιών και την επέκταση της λειτουργικότητας του συστήματος.

- **Μοντέλα πρόβλεψης:** Ένας βασικός τομέας των Big Data Analytics είναι η παραγωγή μοντέλων που να εκτιμούν και να προβλέπουν διάφορες καταστάσεις. Συγκεκριμένα, στον τομέα της υγείας, πρέπει να υπάρχει συνεχής μελέτη των δεδομένων και εκτίμηση σχετικά με αναμενόμενα γεγονότα, ούτως ώστε να μεγιστοποιείται το όφελος και η αξία των δεδομένων.
- **Νομικά ζητήματα:** Υπάρχει πληθώρα ζητημάτων νομικής φύσεως που πρέπει να εξετάζονται και να υπάρχει συνεχής ενημέρωση με τις εξελίξεις στον τομέα αυτό. Χαρακτηριστικά παραδείγματα είναι:
 - **Ασφάλεια δεδομένων:** Οφείλει να εξασφαλίζεται η ασφάλεια του συστήματος από την παράνομη πρόσβαση στα δεδομένα από μη εξουσιοδοτημένα άτομα. Στο πλαίσιο αυτό, οι προκλήσεις ταυτίζονται με αυτές του κλάδου των συστημάτων ασφαλείας, γεγονός που απαιτεί αρκετό χρόνο και προσπάθεια.
 - **Συγκατάθεση ασθενών:** Η συγκατάθεση των ασθενών για κοινή χρήση των προσωπικών δεδομένων είναι μεγάλη πρόκληση, ιδιαίτερα όταν πρόκειται για μεγάλης κλίμακας προσπάθεια. Απαιτείται μεγάλη γραφειοκρατική διαδικασία και χρόνος ώστε να τηρηθούν όλα τα πρωτόκολλα, αλλά είναι κάτι που σίγουρα θα δημιουργήσει ζητήματα σε οργανισμούς υγείας, εάν δεν τηρείται σχολαστικά σε κάθε προσπάθεια αξιοποίησης των Big Data.
 - **Προστασία ιδιωτικότητας ασθενών:** Οι πληροφορίες που αφορούν την υγεία είναι ιδιαίτερα ευαίσθητες και σε καμία περίπτωση δεν πρέπει να υπάρχει δυνατότητα προσδιορισμού, άμεσα ή έμμεσα (βάσει πιθανοτικών μοντέλων), των ατόμων. Ειδικά από τη στιγμή που δεν υφίσταται κάποιο σταθερό και καθολικά αποδεκτό πλαίσιο, η πρόκληση γίνεται ακόμα μεγαλύτερη.



Εικόνα 5.1: Οι κυριότερες προκλήσεις στον τομέα της υγείας σχετικά με την αξιοποίηση των Big Data

Η ανάγκη ανάπτυξης εργαλείων και μεθόδων αντιμετώπισης όλων των προκλήσεων που παρουσιάζονται κατά την αξιοποίηση των Big Data, από οργανισμούς υγείας, απαιτεί συλλογική, οργανωμένη και αυστηρά καθορισμένη προσπάθεια. Παρόλα αυτά, είναι τέτοια η φύση του αντικειμένου, που δεν μπορεί παρά να δοθεί προσοχή σε όλες τις μικρές λεπτομέρειες που προκύπτουν και να διεξαχθεί προσεκτική μελέτη πριν, αλλά και μετά την έναρξη της διαδικασίας.

5.4 Οι νέες τάσεις των Big Data στην υγεία

Υπάρχει μια ισχυρή τάση στην Ιατρική επιστήμη που βασίζεται στην τεκμηρίωση, η οποία περιλαμβάνει τη χρήση όλων των διαθέσιμων κλινικών δεδομένων και τη μετατροπή τους σε προηγμένες αναλύσεις. Η καταγραφή και η συγκέντρωση όλων των πληροφοριών για έναν ασθενή παρέχει μια πληρέστερη εικόνα για την περίθαλψη και τις οικονομικές δραστηριότητες που πραγματοποιούνται, καθώς και συνολικά για τη διαχείριση της υγείας του πληθυσμού. Η εποπτική, ολοκληρωμένη εικόνα του ασθενούς μπορεί επίσης να εξαλείψει τις περιττές δοκιμές και έρευνες που καταναλώνουν σημαντικούς πόρους, οικονομικούς και ανθρώπινους. Είναι δυνατόν επίσης να υπάρξει μείωση στις λανθασμένες συνταγογραφήσεις φαρμάκων και σε αρκετές περιπτώσεις να σωθούν ανθρώπινες ζωές.

Ουσιαστικά, η κυρίαρχη τάση στο χώρο της υγείας που υπόσχεται τις σημαντικότερες καινοτομίες, είναι η βασισμένη στα δεδομένα φροντίδα για των ασθενών. Ο στόχος, λοιπόν, των σύγχρονων συστημάτων υγειονομικής περίθαλψης είναι η παροχή βέλτιστης υγειονομικής περίθαλψης μέσω της χρήσης της τεχνολογίας, προκειμένου να προκύψει:

- Βελτίωση της ποιότητας και του συντονισμού της υγειονομικής περίθαλψης ώστε τα αποτελέσματα να συμβαδίζουν με τις ιατρικές προβλέψεις.
- Μείωση των περιττών εξόδων και της χρήσης των ανθρώπινων πόρων.
- Υποστήριξη σε λειτουργικό και οργανωτικό επίπεδο για τους οργανισμούς υγείας.

Η εστίαση του ενδιαφέροντος της ιατρικής κοινότητας στην παροχή φροντίδας με βάση την αξία, στρέφει την προσοχή στα αποτελέσματα των διαφόρων ιατρικών μεθόδων για την υγεία των ασθενών. Δίνεται μεγάλη σημασία στο βαθμό ικανοποίησης του ασθενούς, ο οποίος είναι πλέον στο επίκεντρο και αυτό επιτυγχάνεται κατά σε μεγάλο βαθμό μέσω των τεχνολογιών των Big Data Analytics.

Οι σημαντικότερες νέες τάσεις στον κλάδο είναι οι ακόλουθες:

Internet of Things (IoT)

Πρόκειται για τον ταχέως αυξανόμενο αριθμό έξυπνων, διασυνδεδεμένων συσκευών και αισθητήρων που παράγουν τεράστιους όγκους δεδομένων. Οι δαπάνες για την υγειονομική περίθαλψη μέσω των IoT παγκοσμίως ξεπέρασαν τα 120 δισεκατομμύρια δολάρια από το 2012 έως το 2016, γεγονός που επιβεβαιώνει τη σημασία τους. Η πλειοψηφία των δεδομένων που δημιουργούνται από τεχνολογίες IoT στον τομέα της υγειονομικής περίθαλψης είναι μη δομημένα, καθιστώντας επιτακτική ανάγκη τη χρήση τεχνολογιών Big Data Analytics για την αξιοποίησή τους.

Ήδη υπάρχει στο εμπόριο πληθώρα συσκευών που παρακολουθούν την ανθρώπινη συμπεριφορά που σχετίζεται με την υγεία, όπως όργανα μέτρησης γλυκόζης, ηλεκτροκαρδιογραφημάτων και αρτηριακής πίεσης. Πολλές από αυτές τις μετρήσεις απαιτούν αξιολόγηση από αρμόδιους ιατρούς. Είναι εξαιρετικής σημασίας η διαπίστωση ότι οι «έξυπνες» συσκευές παρακολούθησης που επικοινωνούν με άλλες συσκευές ασθενών θα μπορούσαν να βελτιώσουν σε μεγάλο βαθμό αυτή τη διαδικασία, πιθανώς ακόμα και να μειώσουν τις ανάγκες άμεσης ιατρικής παρέμβασης σε αρκετές περιπτώσεις. Άλλες έξυπνες συσκευές που υπάρχουν ήδη, μπορούν να ανιχνεύσουν τη συχνότητα χρήσης φαρμάκων στις οικίες των ασθενών. Οι δυνατότητες που προσφέρει η υγειονομική περίθαλψη στον τομέα της υγειονομικής περίθαλψης για τη μείωση του κόστους και τη βελτίωση της περίθαλψης των ασθενών είναι σχεδόν απεριόριστες.

Μείωση της κακοδιαχείρισης και της κατάχρησης πόρων

Το κόστος της κακοδιαχείρισης και της κατάχρησης πόρων στον κλάδο της υγειονομικής περίθαλψης αποτελεί βασικό παράγοντα για την αύξηση των δαπανών ιατροφαρμακευτικής περίθαλψης. Το κλειδί για την αντιμετώπιση του φαινομένου έγκειται στη δυνατότητα αποθήκευσης, επεξεργασίας και ανάλυσης μεγάλων μη δομημένων συνόλων δεδομένων μέσω αλγορίθμων μηχανικής μάθησης οι οποίοι προβλέπουν και αναγνωρίζουν τα «ελλατωματικά» μοτίβα. Οι οργανισμοί υγειονομικής περίθαλψης μπορούν να αναλύουν τα αρχεία των ασθενών και να ανιχνεύουν ανωμαλίες όπως η υπερβολική χρησιμοποίηση των υπηρεσιών του νοσοκομείου σε σύντομες χρονικές περιόδους, εύρεση των ασθενών που λαμβάνουν υπηρεσίες υγειονομικής περίθαλψης από πολλά νοσοκομεία σε διαφορετικές τοποθεσίες ταυτόχρονα ή πολλαπλές συνταγές για τον ίδιο ασθενή που χορηγούνται από διαφορετικούς ιατρούς.

Τήρηση ενιαίων ιατρικών αρχείων των ασθενών

Οι πρωτοβουλίες που σχετίζονται με τα Big Data, επισημαίνουν την ανάγκη για υιοθέτηση συγκεντρωτικών δομών δεδομένων, όπως ο Ηλεκτρονικός Φάκελος Υγείας, ιδιαίτερα λόγω της διαρκούς αύξησης του όγκου της πληροφορίας. Τα σημαντικά ποσά που έχουν διατεθεί για τη δημιουργία και ενημέρωση του Ηλεκτρονικού Φακέλου Υγείας σε μεγάλη κλίμακα, ενισχύουν τον ισχυρισμό αυτό και ουσιαστικά έχουν κοινό στόχο με την αξιοποίηση μεγάλων συνόλων δεδομένων: τη βελτίωση της ακρίβειας της διάγνωσης των ασθενών, την αντιστοίχιση των θεραπειών με τα αντίστοιχα αποτελέσματα και την πρόγνωση παθογενών καταστάσεων. Συνεπώς, η χρήση των Big Data Analytics επισημαίνεται και συμβάλλει άμεσα στην τήρηση των ενιαίων ιατρικών αρχείων των ασθενών.

Περαιτέρω αξιοποίηση τεχνολογιών κινητής και ηλεκτρονικής υγείας

Οι φορείς υγειονομικής περίθαλψης επιδιώκουν να παρέχουν ενισχυμένη προληπτική φροντίδα στους ασθενείς τους με συνεχή παρακολούθηση των ζωτικών τους οργάνων. Τα δεδομένα από την παρακολούθηση μπορούν να αναλυθούν σε πραγματικό χρόνο και να αποσταλούν ειδοποιήσεις στους παρόχους περίθαλψης, ώστε να γνωρίζουν άμεσα τις αλλαγές στην κατάσταση του ασθενούς. Η επεξεργασία γεγονότων σε πραγματικό χρόνο με αλγορίθμους μηχανικής μάθησης μπορεί να προσφέρει στους θεράποντες ιατρούς πληροφορίες για τη λήψη αποφάσεων και να επιτρέψουν την έγκαιρη παρέμβαση, όπου απαιτείται.

Οι αισθητήρες και οι συσκευές δημιουργούν νέες συνθήκες αλληλεπίδρασης με τους ασθενείς με νέους τρόπους, καθιστώντας την υγειονομική περίθαλψη εξαιρετικά πιο αποτελεσματική. Η παρακολούθηση σε πραγματικό χρόνο αλλάζει την ίδια τη φύση της σχέσης ιατρού - ασθενούς. Για παράδειγμα, διατίθενται εφαρμογές που χρησιμοποιούνται για απομακρυσμένη στις οικίες ασθενών με χρόνια αποφρακτική πνευμονοπάθεια. Άλλες εφαρμογές παρακολουθούν το βάρος των ασθενών που καταπολεμούν την αποφρακτική καρδιακή νόσο για την ανίχνευση της κατακράτησης υγρών πριν από τη νοσηλεία. Ορισμένες εφαρμογές καταγράφουν τη χρήση του φαρμάκου για το άσθμα σε παιδιά ώστε να βεβαιωθούν ότι τα μέλη της οικογένειάς τους γνωρίζουν τι πρέπει να διαχειριστεί, περιορίζοντας τις επισκέψεις στα κέντρα υγείας. Προφανώς αυτές οι περιπτώσεις είναι ενδεικτικές και υπάρχει πολύ μεγάλο πλήθος εφαρμογών που καλύπτουν ευρύ φάσμα ασθενειών με ποικίλους τρόπους και συνιστούν τον κλάδο της Κινητής και Ηλεκτρονικής Υγείας (mHealth, eHealth). Ο ρόλος

των Big Data Analytics στο σημείο αυτό είναι κομβικός, καθώς δίνεται η δυνατότητα ομαδοποίησης των στοιχείων από διαφορετικές εφαρμογές και η παραγωγή ιατρικής γνώσης σημαντικά μεγαλύτερης από αυτή στην οποία στοχεύει η κάθε εφαρμογή ανεξάρτητα.

Πέραν λοιπόν των πεδίων που προσφέρονται για εξελίξεις μέσω της τεχνολογίας των Big Data Analytics, και παρουσιάστηκαν στο Κεφάλαιο 4, υπάρχουν και άλλοι τομείς που είναι δυνατόν να αναπτυχθούν. Αυτές οι νέες τάσεις στα Big Data προσθέτουν νέες προοπτικές και συνθέτουν ένα πολύτιμο εργαλείο για την επιστήμη της Ιατρικής και τη βελτίωση των ψηφιακών υπηρεσιών υγείας.

ΚΕΦΑΛΑΙΟ 6

ΕΠΙΛΟΓΟΣ

6.1 Συμπεράσματα

Η παρούσα διπλωματική εργασία καταδεικνύει την πληθώρα δυνατοτήτων για την παροχή πιο στοχευμένης, ευρείας κλίμακας και οικονομικά αποδοτικής υγειονομικής περίθαλψης, μέσω της αξιοποίησης των διαθέσιμων δεδομένων και της χρήσης των Big Data Analytics. Ωστόσο, έχει αποδειχθεί ότι ο τομέας της υγειονομικής περίθαλψης παρουσιάζει συγκεκριμένα χαρακτηριστικά και προκλήσεις που απαιτούν περαιτέρω ερευνητική προσπάθεια, προκειμένου να προκύψει πλήρης αξιοποίηση των δυνατοτήτων που προσφέρονται:

- **Πρόσβαση, διαθεσιμότητα και ποιότητα:** Υπάρχει τεράστιος όγκος δεδομένων που διανέμονται σε πολλά αποθετήρια (silos) και νέα δεδομένα που παράγονται καθημερινά από δισεκατομμύρια συνδεδεμένες συσκευές ή από ανθρώπους. Στο πλαίσιο αυτό, είναι απαραίτητο να εξευρεθούν πιο κατάλληλοι και αποτελεσματικοί τρόποι για τη μόχλευση αυτών των δεδομένων, πάντοτε σε συμφωνία με τις αρχές προστασίας της ιδιωτικών δεδομένων και των δεοντολογικών αρχών, για την πρόσβαση, τη διασφάλιση της ποιότητας και την κατανόηση των σκοπών της χρήσης των Big Data.
- **Ασθενείς και επαγγελματίες υγείας ευνοούνται από την αξιοποίηση των Big Data:** Είναι αδιαμφισβήτητη η ανάγκη ανάπτυξης προσεγγίσεων που να επιτρέπουν σε ανθρώπους και υπολογιστικά συστήματα να συνεργάζονται στενότερα στην εκμετάλλευση των Big Data για καλύτερη περίθαλψη. Αυτό προϋποθέτει την αξιοπιστία των πληροφοριών, την ανάπτυξη κατάλληλων προδιαγραφών και διαδικασιών, τη βελτίωση της διαδραστικότητας και κυρίως την κατανόησης της μεγάλης σημασίας που έχει η επεξεργασία και η ανάλυση των δεδομένων. Οι απαιτήσεις διαφόρων ομάδων ποικίλουν ανάλογα με τους στόχους. Συνεπώς, ερευνητές, ιατροί, νοσηλευτές και ασθενείς απαιτούν διαφορετικές προσεγγίσεις και μεθόδους εστίασης.
- **Περαιτέρω ανάπτυξη τεχνολογικών εργαλείων:** Είναι γεγονός ότι τα τελευταία έτη έχουν δημιουργηθεί πολλά εργαλεία και εμπορικές πλατφόρμες εκμετάλλευσης των Big Data. Παρ' όλα αυτά, απαιτούνται νέες τεχνολογίες που να μπορούν να χειρίζονται, να αναλύουν και να αξιοποιούν αποδοτικότερα το σύνολο των ετερογενών, μη δομημένων και σύνθετων δεδομένων που υπάρχουν ήδη. Βασικότερος στόχος είναι η βελτίωση της καμπύλης εκμάθησης και η μείωση των απαιτήσεων τεχνικών γνώσεων για τη χρήση των εργαλείων. Επίσης, βελτιώσεις στον τρόπο

διαχείρισης και ανάλυσης των δεδομένων είναι δυνατόν να αυξήσουν τις δυνατότητες και να προσφέρουν αποτελέσματα με ακόμα μεγαλύτερη ταχύτητα και αξιοπιστία.

- **Αξιοποίηση της ήδη υπάρχουσας γνώσης:** Παράλληλα με τα διαθέσιμα δεδομένα, υπάρχει τεράστιο πλήθος γνώσης σε βιβλία, επιστημονικά περιοδικά και δημοσιεύσεις. Επίσης, η εμπειρία των επαγγελματιών υγείας είναι πολύτιμη. Για τους λόγους αυτούς απαιτούνται προσεγγίσεις οι οποίες να λαμβάνουν υπόψη και να αξιοποιούν την ήδη υπάρχουσα γνώση. Με τον τρόπο αυτό, θα επιτευχθεί καλύτερη κατανόηση των δεδομένων, ενώ τα δεδομένα θα συμβάλλουν συστηματικά στην παραγωγή νέας γνώσης.
- **Σεβασμός στα προσωπικά δεδομένα:** Για την αξιοποίηση των Big Data στα πλαίσια προστασίας της ιδιωτικότητας, απαιτούνται περαιτέρω πρότυπα και αυστηρά καθορισμένες, ομόφωνα αποδεκτές από τους αρμόδιους φορείς, προδιαγραφές. Οφείλει λοιπόν να υπάρξει σύγκλιση των μοντέλων που αποδέχονται οι φορείς, ούτως ώστε να υπάρχει μοναδική απάντηση όσον αφορά την τήρηση ή μη των απαραίτητων δεοντολογικών αρχών.

6.2 Επίλογος

Είναι πέραν πάσης αμφιβολίας ο τεράστιος αντίκτυπος που μπορεί να επιφέρει η αξιοποίηση της τεχνολογίας των Big Data στην υγεία, τόσο σε επίπεδο βελτίωσης της ποιότητας της περίθαλψης, όσο και στη διαχείριση των μονάδων υγείας, εξοικονομώντας χρόνο, πόρους και παρέχοντας άρτια αποτελέσματα. Ήδη έχει επιβεβαιωθεί η δυνατότητα παροχής εξατομικευμένης ιατρικής φροντίδας. Σηματοδοτείται έτσι μία άτυπη μετάβαση από μοντέλα περίθαλψης που απευθύνονται στο μέσο ασθενή, σε μοντέλα προσαρμοσμένα στον κάθε ασθενή, βάσει του γενετικού του προφίλ, του τρόπου ζωής του, του περιβάλλοντος και άλλων παραγόντων που η παραδοσιακή Ιατρική δεν έχει τη δυνατότητα να συνυπολογίσει, τουλάχιστον σε μεγάλη κλίμακα. Παράλληλα, ο έγκαιρος περιορισμός επιδημιών, η αποτελεσματικότερη διάγνωση και πρόληψη ασθενειών, η εύρεση νέων θεραπευτικών μεθόδων, καθώς και η βελτίωση της ταχύτητας παρασκευής και ποιότητας νέων φαρμάκων, είναι μερικοί ενδεικτικοί παράγοντες που επιβεβαιώνουν ότι η χρήση των Big Data για τη βελτίωση της υγείας έχει εξαιρετικές προοπτικές και θα απασχολήσει την επιστημονική κοινότητα και τους φορείς υγείας στο άμεσο μέλλον. Οι όποιες αμφιβολίες αφορούσαν την είσοδο της τεχνολογίας στην Ιατρική επιστήμη έχουν ήδη καμφθεί, γεγονός που ευνοεί την αποδοχή της ευρείας χρήσης τους.

Δεν τίθεται αμφιβολία ότι οι επενδύσεις σε οικονομικούς και ανθρώπινους πόρους για τη βελτίωση των υπηρεσιών υγείας μέσω των Big Data Analytics θα αυξηθούν στο άμεσο μέλλον. Το πλήθος των προβλημάτων που επιλύονται μέσω της χρήσης τους είναι τεράστιο, ενώ προς το παρόν δεν φαίνεται να υπάρχει κάποια εναλλακτική τεχνολογία με συγκρίσιμες προοπτικές. Για το λόγο αυτό, κυριαρχεί η εκτίμηση πως η αξιοποίηση των δεδομένων σε μεγάλη κλίμακα δεν θα απασχολεί στο μέλλον μόνο τους «μεγάλους» φορείς και οργανισμούς, αλλά θα πρέπει κάθε κλινική και κάθε ιατρός να αξιοποιεί τα διαθέσιμα τεχνολογικά εργαλεία με στόχο τη βέλτιστη παροχή υπηρεσιών υγείας. Αναπόφευκτη είναι η δημιουργία νέων θέσεων απασχόλησης και η ζήτηση ατόμων με τεχνικές γνώσεις, καθώς είναι σημαντικά τα χρηματικά ποσά που δαπανώνται άσκοπα, είτε λόγω αναποτελεσματικής διαχείρισης που οφείλεται σε λανθασμένους χειρισμούς, είτε λόγω λανθασμένων θεραπειών και διαγνώσεων. Κυρίως, όμως ο ανθρώπινος παράγοντας, δηλαδή η ριζική αναβάθμιση των υπηρεσιών υγείας που εγκαινιάζει μία νέα εποχή, είναι ο σημαντικότερος λόγος που καταργεί κάθε αμφιβολία για την διάδοση των Big Data Analytics στο άμεσο μέλλον.

Το μέγεθος της συνεισφοράς είναι συνάρτηση πολλών παραγόντων και αδύνατο να εκτιμηθεί με ακρίβεια. Αυτό οφείλεται στο χρόνο που απαιτείται ώστε κάποια καινοτομία να γνωρίσει ευρεία αποδοχή. Πέραν λοιπόν της ενημέρωσης των επιστημόνων υγείας, επιβάλλεται η ανάληψη πολιτικής πρωτοβουλίας σε εθνικό και διεθνές επίπεδο, η δημιουργία ισχυρότερου κινήτρου για τους οργανισμούς και επαγγελματίες υγείας για την αξιοποίηση της τεχνολογίας αυτής, καθώς και η εκπαίδευση ανθρώπινου δυναμικού με τις κατάλληλες τεχνικές και οργανωτικές δεξιότητες ούτως ώστε να δομηθεί το εγχείρημα εξ αρχής σε σταθερές βάσεις. Η εκτίμηση για το χρονικό ορίζοντα στον οποίο τα Big Data Analytics στην ψηφιακή υγεία θα θεωρούνται όχι μία νέα τεχνολογία, αλλά ένα απαραίτητο εφόδιο για τη φροντίδα των ασθενών είναι εξαιρετικά δύσκολο να πραγματοποιηθεί. Το μόνο βέβαιο είναι πως το μέλλον εξαρτάται αποκλειστικά από τον ανθρώπινο παράγοντα και πέραν των εξελίξεων που έχουν ήδη σηματοδοτηθεί, τα μεγάλα δεδομένα αναμένεται να αποτελέσουν εφαλτήριο νέων ριζικών αλλαγών και καινοτομιών στον τομέα της υγείας.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Raghupathi N., Raghupathi V., «Big Data analytics in healthcare: promise and potential», *Health Information Science and Systems*, 2014; 2:3
2. John R. Mashey (25 April 1998). "Big Data ... and the Next Wave of InfraStress" (PDF). Slides from invited talk. Usenix.
3. Steve Lohr (1 February 2013). "The Origins of 'Big Data': An Etymological Detective Story". *New York Times*.
4. Crawford, Kate (September 21, 2011). "Six Provocations for Big Data". *Social Science Research Network: A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*.
5. Jacobs, A. (6 July 2009). "The Pathologies of Big Data". *ACMQueue*.
6. Hilbert, Martin. "Big Data for Development: A Review of Promises and Challenges. *Development Policy Review*". *martinhilbert.net*.
7. https://www.ghdonline.org/uploads/big-data-in-healthcare_B_Kaplan_2012.pdf. Retrieved at 25 April, 2017
8. <http://www-935.ibm.com/services/us/gbs/thoughtleadership/ibv-healthcare-analytics.html> . Retrieved at 25 April, 2017
9. <https://www.ibm.com/developerworks/wikis/display/db2oncampus/FREE+ebook+-+Understanding+Big+Data> Retrieved at 25 April, 2017
10. Tutorial on semi-structured data by Peter Buneman from *Symposium on Principles of Database Systems*, 1997 [1]
11. Frost & Sullivan, "U.S. Hospital Health Data Analytics Market: Growing EHR Adoption Fuels A New Era in Analytics", 2012.
12. Y.Koumpouros, "Big Data in Healthcare" in *Cloud Computing Applications for Quality Health Care Delivery*, IGI Global Publishing, 2014, ISBN: 978-1-4666-6118-9
13. <https://hbr.org/2011/09/how-to-solve-the-cost-crisis-in-health-care> Retrieved at 25 April, 2017
14. <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation> . Retrieved at 25 April, 2017
15. http://www.kauffman.org/~media/kauffman_org/research%20reports%20and%20covers/2012/04/valuing_health_care.pdf . Retrieved at 25 April, 2017

16. Γιάννης Κυριόπουλος. Περιοδικό “ΧΡΟΝΟΣ” // τεύχος ΤΡΙΑΝΤΑ ΠΕΝΤΕ, Μάρτιος. 2016
17. Bertot J.C., Jaeger P., Hansen D. «The impact of polices on government social media usage: Issues, challenges, and recommendations.» *Government Information Quarterly*, 2012: 29 (1); 30-40 doi:10.1016/j.giq.2011.04.004
18. ["Welcome to Apache Hadoop!"](http://hadoop.apache.org/). *hadoop.apache.org*. [Accessed: 05 April 2017]
19. ["What is the Hadoop Distributed File System \(HDFS\)?"](http://ibm.com/IBM). *ibm.com*. *IBM*. Retrieved 2014-10-30.
20. Malak, Michael (2014-09-19). ["Data Locality: HPC vs. Hadoop vs. Spark"](http://datascienceassn.org). *datascienceassn.org*. Data Science Association. [Accessed: 05 April 2017].
21. ["Hadoop-related projects at"](http://hadoop.apache.org/). *Hadoop.apache.org*. [Accessed: 05 April 2017].
22. Zaharia, Matei; Chowdhury, Mosharaf; Franklin, Michael J.; Shenker, Scott; Stoica, Ion. [Spark: Cluster Computing with Working Sets](#) (PDF). *USENIX Workshop on Hot Topics in Cloud Computing (HotCloud)*.
23. Harris, Derrick (28 June 2014). ["4 reasons why Spark could jolt Hadoop into hyperdrive"](#). *Gigaom*.
24. <https://www.talend.com/> Retrieved 24 April, 2017
25. IBM, “InfoSphere Platform”, 2017. [Online]. Available: <http://www01.ibm.com/software/data/infosphere/> . [Accessed: 05 April 2017].
26. <https://datacleaner.org/> . Retrieved 24 April, 2017
27. [«The RedMonk Programming Language Rankings: January 2013 – tecosystems»](#). Retrieved 24 April, 2017
28. <https://www.r-project.org/about.html> . Retrieved 24 April, 2017
29. Martin Odersky et al., *An Overview of the Scala Programming Language, 2nd Edition*
30. Deitel, Paul. Deitel, Harvey (2015). *Java Προγραμματισμός (Ελληνική Μετάφραση)* (10η έκδοση). Αθήνα: Μ. Γκιούρδας, σελ. 19. [ISBN 978-960-512-681-0](#).
31. *Aws.amazon.com*, “Amazon Web Services, Inc. Homepage”, 2017. [Online]. Available: <http://aws.amazon.com>. [Accessed: 05 April 2017].
32. *Cloudera.com*, “Cloudera, Inc. Homepage”, 2017. [Online]. Available: www.cloudera.com. [Accessed: 05 April 2017].

33. *Datameer.com* , “*Datameer, Inc. Homepage*” 2017. [Online]. Available: www.datameer.com. [Accessed: 07 April 2017].
34. *Investing in Health*: http://ec.europa.eu/health/strategy/docs/swd_investing_in_health.pdf . Retrieved 14 April, 2017
35. <http://www.oecd.org/eco/growth/46508904.pdf> . Retrieved 14 April, 2017
36. <http://www.nextech.com/blog/healthcare-data-growth-exponential-problem> . Retrieved 14 April, 2017
37. http://www.healthparliament.eu/documents/10184/0/EHP_papers_BIGDATAINHEALTHCARE.pdf . Retrieved 14 April, 2017
38. <https://www.healthdatamanagement.com/news/how-data-pros-are-trying-to-achieve-roi-with-big-data/> . Retrieved 24 April, 2017
39. <http://www.euro.who.int/en/health-topics/Life-stages/healthy-ageing/data-and-statistics> . Retrieved 14 April, 2017
40. <http://www.bdva.eu/sites/default/files/Big%20Data%20Technologies%20in%20Healthcare.pdf> . Retrieved 14 April, 2017
41. http://www.euro.who.int/_data/assets/pdf_file/0008/96632/E93736.pdf . Retrieved 14 April, 2017
42. Kissick, W. (1994). “*Medicine’s Dilemmas*”. New Haven and New London, CT: Yale University Press
43. <https://www.idc.com/> . Retrieved 24 April, 2017
44. <https://www.information-management.com/news/big-data-and-analytics-see-double-digit-growth-through-2020> . Retrieved 24 April, 2017
45. <https://www.marketsandmarkets.com> , “*Healthcare Analytics/Medical Analytics Market by Application (Clinical, PHM, Financial (RCM, Claim & Fraud), Supply Chain & HR), Type (Predictive), Delivery model (On-premise, Cloud), End-user (Payer, Hospital, Ambulatory, ACO) - Global Forecast to 2020*”, July 2015
46. <https://www.slideshare.net/AndersQuitzaulbm/ibm-watson-in-healthcare> . Retrieved 24 April, 2017

47. C. A. McGinn, S. Grenier, J. Duplantie, N. Shaw, C. Sicotte, L. Mathieu, Y. Leduc, F. Légaré, and M.- P. Gagnon, "Comparison of user groups' perspectives of barriers and facilitators to implementing electronic health records: a systematic review.," *BMC Med.*, vol. 9, p. 46, 2011
48. *Health Information Systems – Past, Present, Future*. Reinhold Haux. Technical University of Braunschweig, Institute for Medical Informatics, Muehlenpfordtstr. *International Journal of Medical Informatics* (2006) 75, 268—281
49. *Percentage of Physicians with Electronic Health Record (EHR) Systems That Meet Federal Standards,* by Physician Specialty — Physician Workflow Survey, United States, 2011*
50. *Big Data and Analytics for Infectious Disease Research, Operations, and Policy: Proceedings of a Workshop, 2016 17 Shiha*
51. Shihao Yang, Mauricio Santillana, and S. C. Kou *Accurate estimation of influenza epidemics using Google search data via ARGO PNAS 2015*, <http://www.pnas.org/content/112/47/14473>
52. <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0019467> . Retrieved 24 April, 2017
53. Porter, M. E., & Teisberg, E. O. (2006). *Redefining health care: creating value-based competition on results*. Harvard Business Press.
54. *European Centre for Disease Prevention and Control*:
http://ecdc.europa.eu/en/healthtopics/Healthcareassociated_infections/Pages/index.aspx . Retrieved 24 April, 2017
55. *HAI Data and Statistics, Centers for Disease Control and Prevention (2016)*
<http://www.cdc.gov/HAI/surveillance/> . Retrieved 24 April, 2017
56. [https://en.wikipedia.org/wiki/Watson_\(computer\)#cite_note-atlantic20110217-10](https://en.wikipedia.org/wiki/Watson_(computer)#cite_note-atlantic20110217-10) . Retrieved 24 April, 2017
57. Wakeman, Nick (February 17, 2011). "[IBM's Watson heads to medical school](#)". *Washington Technology*. Retrieved 14 April, 2017
58. Mathews, Anna Wilde (September 12, 2011). "[Wellpoint's New Hire: What is Watson?](#)". *The Wall Street Journal*
59. Miliard, Mike (October 30, 2012). "[Watson Heads to Medical School: Cleveland Clinic, IBM Send Supercomputer to College](#)". *Healthcare IT News*. Retrieved 14 April, 2017

60. ["MD Anderson Taps IBM Watson to Power "Moon Shots" Mission Aimed at Ending Cancer, Starting with Leukemia"](#) (Press release). IBM.
61. ["IBM's Watson Now Tackles Clinical Trials At MD Anderson Cancer Center"](#). Forbes.
62. ["MD Anderson Benches IBM Watson In Setback For Artificial Intelligence In Medicine"](#). Forbes
63. Leske, Nikola (February 9, 2013). ["Doctors Seek Help on Cancer Treatment from IBM Supercomputer"](#). Reuters. Retrieved 14 April, 2017
64. ANI (2016-10-28). ["Manipal Hospitals to adopt IBM's 'Watson for Oncology' supercomputer for cancer treatment"](#). Business Standard India. Retrieved 14 April, 2017
65. McCurry, Justin (2017-01-05). ["Japanese company replaces office workers with artificial intelligence"](#). The Guardian. ISSN 0261-3077. Retrieved 14 April, 2017
66. <https://www-03.ibm.com/press/us/en/pressrelease/46748.wss> Available: www.ibm.com. . Retrieved 14 April, 2017
67. <http://www.medicaldaily.com/5-ways-ibm-watson-changing-health-care-diagnosing-disease-treating-it-364394> . Retrieved 14 April, 2017
68. <http://www.healthcarediver.com/news/ibm-unveils-new-imaging-solutions-from-watson-health-merge-healthcare/431406/> . Retrieved 14 April, 2017
69. <https://globalgenes.org/rare-diseases-facts-statistics/> . Retrieved 14 April, 2017
70. <http://data.worldbank.org/indicator/SH.MED.PHYS.ZS> . Retrieved 14 April, 2017
71. <http://www.prnewswire.com/news-releases/manipal-hospitals-adopts-watson-for-oncology-to-help-physicians-identify-options-for-individualized-evidence-based-cancer-care-across-india-300186824.html> , 2017. [Online]. Available: www.prnewswire.com. Retrieved 14 April, 2017
72. <http://www.medicaldaily.com/ibms-watson-guide-cancer-therapies-14-centers-332140>. Retrieved 14 April, 2017
73. <http://fortune.com/2015/12/10/ibm-watson-nutrition/> . Retrieved 14 April, 2017
74. <https://obamawhitehouse.archives.gov/the-press-office/2015/01/30/fact-sheet-president-obamas-precision-medicine-initiative> . Retrieved 14 April, 2017
75. http://www.emif.eu/assets/e/m/emif_summary_till_end_2015_website.pdf . Retrieved 14 April, 2017

76. <http://omop.org/CDM> Retrieved 14 April, 2017
77. <http://www.emif.eu/results> Retrieved 14 April, 2017
78. [Williams, A. J.](#); Harland, L.; Groth, P.; [Pettifer, S.](#); Chichester, C.; Willighagen, E. L.; Evelo, C. T.; Blomberg, N.; [Ecker, G.](#); [Goble, C.](#); [Mons, B.](#) (2012). "Open PHACTS: Semantic interoperability for drug discovery". *Drug Discovery Today*. 17 (21–22): 1188–1198. [doi:10.1016/j.drudis.2012.05.016](#). [PMID 22683805](#).
79. [Open PHACTS at IMI - The Innovative Medicines Initiative](#)". Retrieved 14 April, 2017
80. [Goble, C.](#); Gray, A. J. G.; Harland, L.; Karapetyan, K.; Loizou, A.; Mikhailov, I.; Rankka, Y. N.; Senger, S.; Tkachenko, V.; [Williams, A. J.](#); Willighagen, E. L. (2013). "Incorporating Commercial and Private Data into an Open Linked Data Platform for Drug Discovery". *The Semantic Web – ISWC 2013. Lecture Notes in Computer Science*. 8219. p. 65. [doi:10.1007/978-3-642-41338-4_5](#). [ISBN 978-3-642-41337-7](#).
81. [Open PHACTS on GitHub](#)". Retrieved 19 April, 2017
82. <https://dev.openphacts.org/> . Retrieved 19 April, 2017
83. <http://www.nature.com/news/first-paralysed-person-to-be-reanimated-offers-neuroscience-insights-1.19749> . Retrieved 24 April, 2017
84. *The Feasibility of a Brain-Computer Interface Functional Electrical Stimulation System for the Restoration of Overground Walking after Paraplegia* Christine E King, Po T Wang, Colin M McCrimmon, Cathy CY Chou, An H Do, Zoran Nenadic *Journal of NeuroEngineering and Rehabilitation* 2015
85. *Big Data research: Artemis, Apollo and Athena* Dr. Carolyn McGregor AM Canada Research Chair in Health Informatics University of Ontario Institute of Technology
86. *Making Good Prediction: A Theoretical Framework*. Adeline Lo, Herman Chernoff, Tian Zheng and Shaw-Hwa Lo. October 2015
87. *Anticipating the international spread of Zika virus from Brazil* Isaac I. Bogoch, Oliver J. Brady, Moritz U. G. Kramer, Matthew German, Marisa I. Creatore, Manisha A. Kulkarni, John S. Brownstein, Sumiko R. Mekaru, Simon I. Hay, Emily Groot, Alexander Watts, Kamran Khan. Retrieved at [http://thelancet.com/journals/lancet/article/PIIS0140-6736\(16\)00080-5/fulltext](http://thelancet.com/journals/lancet/article/PIIS0140-6736(16)00080-5/fulltext) . 25 April, 2017

88. <https://www.forbes.com/sites/davidshaywitz/2015/03/24/data-silos-healthcares-silent-tragedy/#1ec175372907> . Retrieved 19 April, 2017
89. Crawford K., Schulz J., «Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms». *Boston College Law Review* 2014: 55; 93:1-4.
90. Nagar R, Yuan Q, Freifeld CC, Santillana M, Nojima A, Chunara R, Brownstein JS «A Case Study of the New York City 2012-2013 Influenza Season With Daily Geocoded Twitter Data From Temporal and Spatiotemporal Perspectives». *J Med Internet Res* 2014;16(10):e236, DOI: 10.2196/jmir.3416
91. Chon J., Raymond R., Wang H., Wang F «Modeling Flu Trends with Real-Time Geo-tagged Twitter Data Streams» In: Xu K. , Zhu J. (Eds) *Wireless Algorithms, Systems, and Applications*, 2015; LCNS 9204: 60-69 DOI: 10.1007/978-3-319-21837-3 7
92. Dove, E.S., Özdemir V. «What Role for Law, Human Rights, and Bioethics in an Age of Big Data, Consortia Science, and Consortia Ethics? The Importance of Trustworthiness». *Laws* 2015; 4(3): 515-540. doi:10.3390/laws4030515
93. <https://assets.sourcemia.com/a0/c6/296701f34ab69e369dd9c9c1345f/unlocking-big-data-for-healthcare-white-paper.pdf>
94. <http://www.investopedia.com/terms/s/safeharbor.asp> . Retrieved 19 April, 2017
95. <http://www.wipo.int/amc/en/expert-determination/what-is-exp.html> / . Retrieved 19 April, 2017
96. <https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/> . Retrieved 19 April, 2017
97. Hall, Susan D. (2016, February 17). *ONC: Patient comfort levels with EHRs, data-sharing on the rise*. *FierceHealthIT*. Retrieved from <http://www.fiercehealthcare.com/it/onc-patient-comfort-levels-ehrs-datasharing-rise> . Retrieved 19 April, 2017
98. <https://assets.sourcemia.com/a0/c6/296701f34ab69e369dd9c9c1345f/unlocking-big-data-for-healthcare-white-paper.pdf> . Retrieved 19 April, 2017