

ΠΑΡΑΔΟΤΕΟ Π12.2

Σχεδιασμός και Ανάπτυξη Ευφυούς Μηχανισμού Δημιουργίας Νέων Μητρώων Ασθενών

Intelligent Patient Registry Creation Mechanism

Κωδικός Πράξης	TAEDR-0539180
Τίτλος Έργου	Γεφυρώνοντας μεγάλα ομικά, γενετικά και ιατρικά δεδομένα για την ευρεία εφαρμογή της Ιατρικής Ακριβείας στην Ελλάδα
Ενότητα Εργασίας	EE12: Ανάπτυξη ολοκληρωμένου υπολογιστικού συστήματος Έξυπνου Ηλεκτρονικού Φακέλου Υγείας
Είδος Παραδοτέου	Λογισμικό
Υπεύθυνος Φορέας	Πανεπιστήμιο Θεσσαλίας
Συμμετέχοντες Φορείς	Πανεπιστήμιο Θεσσαλίας
Ημερομηνία Υποβολής	31 Δεκεμβρίου 2025
Έκδοση	2.0



Με τη χρηματοδότηση
της Ευρωπαϊκής Ένωσης
NextGenerationEU

Ελλάδα 2.0 - Εθνικό Σχέδιο Ανάκαμψης και Ανθεκτικότητας
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Πίνακας Περιεχομένων

Περιεχόμενα

Πίνακας Περιεχομένων	3
Πίνακας Πινάκων.....	7
Πίνακας Σχημάτων.....	8
Περίληψη	9
1. Εισαγωγή: Μητρώα Ασθενών στην Ιατρική Ακριβείας	10
1.1 Ορισμός και Έννοια των Μητρώων Ασθενών	10
1.2 Ο Ρόλος των Μητρώων στο Σύγχρονο Σύστημα Υγείας.....	10
1.2.1 Έρευνα και Επιδημιολογία.....	10
1.2.2 Analytics και Real World Evidence.....	10
1.2.3 Πολιτικές Υγείας και Health Technology Assessment	11
1.3 Παραδείγματα Μητρώων Ασθενών	11
1.3.1 Μητρώα στην Ελλάδα	11
1.3.2 Διεθνή Παραδείγματα.....	11
1.4 Η Ανάγκη για Μητρώα με Γενετικά Δεδομένα	12
1.4.1 Περιορισμοί των Παραδοσιακών Μητρώων	12
1.4.2 Μητρώα στην Εποχή της Ιατρικής Ακριβείας.....	12
1.4.3 Στόχοι του Παρόντος Παραδοτέου	12
2. Αρχιτεκτονική Συστήματος Μητρώων	13
2.1 Σχεδιαστικές Αρχές	13
2.2 Αρχιτεκτονική Δύο Επιπέδων	13
2.2.1 Επιχειρησιακό Επίπεδο (FHIR R4)	13
2.2.2 Αναλυτικό Επίπεδο (OMOP CDM).....	14
2.2.3 Γιατί Χρειάζονται και τα Δύο Επίπεδα	14
2.3 Δεδομένα που Καταγράφονται.....	15
2.3.1 Κλινικά Δεδομένα.....	15
2.3.2 Γενετικά Δεδομένα.....	15
2.4 OMOP CDM και Genomic CDM Extension.....	15
2.4.1 Βασικοί Πίνακες OMOP	15
2.4.2 Genomic CDM Extension	16
2.4.3 Mapping FHIR προς OMOP.....	16
2.5 Διαλειτουργικότητα.....	17
2.5.1 FHIR APIs για Ανταλλαγή Δεδομένων.....	17
2.5.2 OMOP για Cross-Institutional Analytics.....	17
2.5.3 Συμμόρφωση με EHDS	17

3. Κύκλος Ζωής Μητρώου	18
3.1 Ορισμός Μητρώου (Registry Definition Phase)	18
3.1.1 Ορισμός Κοόρτης (Cohort Definition)	18
3.1.2 Προδιαγραφές Dataset (Dataset Specifications)	19
3.1.3 Δείκτες Ποιότητας και Αποτελεσμάτων (Quality and Outcome Indicators)	20
3.1.4 Διακυβέρνηση Μητρώου (Registry Governance)	21
3.2 Μηχανισμοί Εμπλουτισμού Δεδομένων (Data Enrichment Mechanisms)	21
3.2.1 Event-Driven (Live) Enrichment	21
3.2.2 Batch ETL (Extract-Transform-Load)	22
3.2.3 Federated Approach (Ομοσπονδιακή Προσέγγιση)	23
3.3 Αυτόματος έναντι Χειροκίνητου Εμπλουτισμού	23
3.3.1 Αυτόματος Εμπλουτισμός	23
3.3.2 Χειροκίνητος Εμπλουτισμός	24
3.4 Διασφάλιση Ποιότητας Δεδομένων (Data Quality Assurance)	24
3.4.1 Κανόνες Επικύρωσης (Validation Rules)	24
3.4.2 Παρακολούθηση Ποιότητας (Quality Monitoring)	25
3.5 Εξέλιξη Μητρώου και Διαχείριση Εκδόσεων	25
3.5.1 Εξέλιξη Σχήματος Δεδομένων	25
3.5.2 Διαδικασία Διαχείρισης Αλλαγών	25
4. Ασφάλεια και Συμμόρφωση	26
4.1 Συμμόρφωση με GDPR	26
4.1.1 Νομική Βάση Επεξεργασίας	26
4.1.2 Ελαχιστοποίηση Δεδομένων	26
4.1.3 Δικαιώματα Υποκειμένων	26
4.2 Έλεγχος Πρόσβασης	26
4.2.1 Role-Based Access Control (RBAC)	26
4.2.2 Διαχείριση Συγκαταθέσεων	26
4.3 Audit και Ιχνηλασιμότητα	27
4.3.1 Audit Trail	27
4.3.2 Data Provenance	27
5. Τεχνική Υλοποίηση	28
5.1 Registry Engine	28
5.1.1 Σύστημα Επισήμανσης Μητρώων (Registry Tagging)	28
5.1.2 Αρχιτεκτονική ETL Pipeline	28
5.1.3 Cohort Builder	29
5.1.4 Scheduler και Batch Processing	29
5.2 Registry Store	29
5.2.1 FHIR Server (Operational Layer)	29
5.2.2 OMOP Database (Analytics Layer)	29

5.2.3 ETL Pipeline FHIR → OMOP.....	30
5.3 Έλεγχοι Ποιότητας Δεδομένων.....	30
5.3.1 Validation Rules	30
5.3.2 Completeness Checks	30
5.3.3 Consistency Checks.....	30
5.3.4 Reconciliation με Κλινικούς.....	31
5.4 Dashboards και Αναφορές.....	31
5.4.1 Registry Overview Dashboard	31
5.4.2 Clinical Indicators	31
5.4.3 Export Functionality.....	31
5.5 Σύνοψη Κατάστασης Υλοποίησης.....	31
6. Υλοποίηση στα Πλαίσια του Έργου.....	32
6.1 Κοινό Μοντέλο Δεδομένων FHIR.....	32
6.1.1 Μοντελοποίηση Ασθενών (Patient Resource).....	32
6.1.2 Μοντελοποίηση Γενετικών Ευρημάτων	32
6.1.3 Κλινική Σημασία και Ταξινόμηση ACMG	33
6.1.4 Ειδικές Επεκτάσεις για τον Ελληνικό Πληθυσμό	34
6.2 WES Registry - Μητρώο Σπάνιων Νοσημάτων.....	35
6.2.1 Αρχιτεκτονική FHIR Resources	36
6.2.2 Πηγή και Μορφή Δεδομένων	36
6.2.3 Πρότυποι Κωδικοί LOINC.....	37
6.2.4 Ταξινόμηση Κλινικής Σημαντικότητας (ACMG)	38
6.2.5 Ενσωμάτωση HPO (Human Phenotype Ontology)	38
6.2.6 Ταξινόμηση Τύπων Παραλλαγών	39
6.2.7 Κατηγοριοποίηση Νοσημάτων κατά ICD-10.....	39
6.2.8 Χειρισμός Σύνθετης Ετεροζυγωτίας	40
6.2.9 Ζυγωτία και Πρότυπα Κληρονομικότητας	40
6.2.10 Αγωγός Δεδομένων (Data Pipeline).....	40
6.2.11 Σύστημα Tags για Απομόνωση Πόρων	41
6.2.12 Στατιστικά και Δείκτες Μητρώου.....	41
6.3 CanVaS Registry - Μητρώο Κληρονομικού Καρκίνου	42
6.3.1 Προέλευση και Δομή Δεδομένων.....	43
Μεθοδολογία Γενετικού Ελέγχου.....	43
6.3.2 Γονίδια Προδιάθεσης στον Καρκίνο.....	43
Κατηγοριοποίηση Γονιδίων κατά Διεισδυτικότητα:	43
Κύρια Κλινικά Σύνδρομα και Σχετιζόμενα Γονίδια:.....	44
6.3.3 Ελληνικές Founder Mutations.....	44
6.3.4 Ενσωμάτωση Φαινοτυπικών Δεδομένων.....	45
6.3.5 Τεχνική Ενσωμάτωση στο FHIR.....	46

6.3.6 Κλινική Αξιοποίηση και Μελλοντικές Προοπτικές	47
6.4 OMOP CDM και Αναλυτικό Επίπεδο	48
6.4.1 Διαδικασία ETL FHIR → OMOP	48
6.4.2 Genomic CDM Extension Tables	49
6.4.3 Federated Analytics Capability	50
6.4.4 Decentralized Analysis Workflows	50
6.5 Επαναρησιμοποιήσιμα Στοιχεία	51
6.5.1 ETL Pipeline Components	51
6.5.2 Cohort Definition Library	51
7. Κλιμάκωση και Επεκτασιμότητα	52
7.1 Επέκταση σε Νέα Μητρώα.....	52
7.1.1 Template-Based Creation.....	52
7.1.2 Self-Service για Κλινικούς Ερευνητές	52
7.1.3 Διαδικασία Δημιουργίας Νέου Μητρώου.....	52
7.2 Συμμετοχή σε Federated Δίκτυα.....	53
7.2.1 Σύνδεση με το European Health Data Space (EHDS)	53
7.2.2 Συμμετοχή στο Δίκτυο OHDSI.....	53
7.2.3 Cross-Border Genomic Analytics.....	54
7.2.4 Αρχιτεκτονική Federation Node.....	54
7.3 Τεχνική Κλιμάκωση	54
7.3.1 Στρατηγικές Κλιμάκωσης	54
7.3.2 Εκτιμώμενη Χωρητικότητα.....	55
7.4 Μελλοντικές Επεκτάσεις.....	55
7.4.1 Βραχυπρόθεσμες Επεκτάσεις	55
7.4.2 Μεσοπρόθεσμες Επεκτάσεις	55
7.4.3 Μακροπρόθεσμες Επεκτάσεις	55
8. Συμπεράσματα	57
8.1 Σύνοψη Επιτευγμάτων	57
8.2 Καινοτομία.....	57
8.3 Συμβολή στους Στόχους του Έργου.....	57
8.4 Μελλοντική Ανάπτυξη.....	58
9. Αναφορές	59
9.1 Πρότυπα και Specifications	59
9.2 Μητρώα και Δίκτυα.....	59
9.3 Coding Systems	60
9.4 Γενετική και Ταξινόμηση Παραλλαγών	60
Παράρτημα Α: Γλωσσάρι.....	61
Παράρτημα Β: Πίνακες Genomic CDM.....	62

Πίνακας Πινάκων

Πίνακας 1 FHIR σε OMOP mapping.....	17
Πίνακας 2 Παραδείγματα Ορισμών Κούρτης	19
Πίνακας 3 Κρισιμότητα και Κατηγοριοποίηση.....	19
Πίνακας 4 Διαλειτουργικότητα και Συστήματα Κωδικοποίησης.....	20
Πίνακας 5 Παρακολούθηση ποιότητας Δεδομένων	20
Πίνακας 6 Δείκτες και αποτελέσματα.....	20
Πίνακας 7 Ρόλοι Διακυβέρνησης Μητρώου	21
Πίνακας 8 Περιπτώσεις Χρήσης Event-Driven Enrichment	22
Πίνακας 9 Προγραμματισμός ETL Jobs	22
Πίνακας 10 Σύγκριση Κεντρικής και Federated Προσέγγισης.....	23
Πίνακας 11 Σύγκριση Μεθόδων Εμπλουτισμού	24
Πίνακας 12 Κανόνες και κατηγορίες ελέγχου.....	24
Πίνακας 13 Ρόλοι και Πρόσβαση	26
Πίνακας 14 Μητρώο και Tagging.....	28
Πίνακας 15 Μητρώο και Δεδομένα	29
Πίνακας 16 Αντιστοίχιση FHIR Resources σε OMOP Tables	30
Πίνακας 17 Κατάσταση Υλοποίησης.....	31
Πίνακας 18 Πεδία και Μοντελοποίηση.....	32
Πίνακας 19 Gene Observation - Βασικά στοιχεία.....	33
Πίνακας 20 Variant Observation - Βασικά στοιχεία	33
Πίνακας 21 ACMG και Κλινική Σημασία	33
Πίνακας 22 Κανόνες και Ερμηνεία.....	34
Πίνακας 23 Γονίδια και Παραλλαγές.....	34
Πίνακας 24 Γεωγραφικές Περιοχές Παραλλαγών.....	35
Πίνακας 25 Ιεραρχία FHIR Resources του WES Registry.....	36
Πίνακας 26 Κατάλογος Σηλών Πηγής Δεδομένων WES	37
Πίνακας 27 Κωδικοί LOINC για Γενετικές Παρατηρήσεις	37
Πίνακας 28 Αντιστοίχιση Ταξινομήσεων ACMG.....	38
Πίνακας 29 Συνθήκη και Αποτελέσματα	38
Πίνακας 30 Περιγραφή Συμπτώματος	38
Πίνακας 31 Αντιστοίχιση Τύπων Παραλλαγών	39
Πίνακας 32 Κατηγορίες ICD-10 στο WES Registry	39
Πίνακας 33 Αντιστοίχιση Ζυγωτίας.....	40
Πίνακας 34 Αντιστοίχιση Προτύπων Κληρονομικότητας.....	40
Πίνακας 35 Πόροι και Εξαρτήσεις.....	41
Πίνακας 36 Σύστημα Tags WES Registry	41
Πίνακας 37 Μετρικές και Τιμές του CanVas	43
Πίνακας 38 Παραδείγματα Κατηγοριοποίησης Γονιδίων	43
Πίνακας 39 Συσχέτιση σύνδρομου και γονιδίου	44
Πίνακας 40 Τεκμηριωμένες Ελληνικές Founder Mutations.....	44
Πίνακας 41 Κατανομή Φαινοτύπων στο CanVaS.....	45
Πίνακας 42 FHIR Resource Mapping ανά Ασθενή.....	46
Πίνακας 43 URL Extensions.....	46
Πίνακας 44 Identifier System	47
Πίνακας 45 Tag System.....	47
Πίνακας 46 Αντιστοίχιση FHIR Resources σε OMOP Tables	48
Πίνακας 47 Γονίδια προς Εξέταση.....	49
Πίνακας 48 Πεδία και Παραλλαγές	49

Πίνακας 49 Εμνηνεία Παραλλαγών.....	50
Πίνακας 50 Πλεονεκτήματα Τεχνολογίας Federated	50
Πίνακας 51 ETL Components.....	51
Πίνακας 52 Cohort Definition.....	51
Πίνακας 53 Διαδικασία Δημιουργίας Μητρώου	52
Πίνακας 54 Συμμόρφωση με Απαιτήσεις EHDS	53
Πίνακας 55 Εργαλεία OHDSI για Federated Analytics	53
Πίνακας 56 Στρατηγικές Κλιμάκωσης.....	54
Πίνακας 57 Εκτιμήσεις Χωρητικότητας Συστήματος.....	55

Πίνακας Σχημάτων

Σχήμα 1 Αρχιτεκτονική Smart EHR Platform	13
Σχήμα 2 Ροή μετάβασης δεδομένων από το FHIR στο OMOP.....	14
Σχήμα 3 Αποθετήριο για την ανάπτυξη του genomic module - CDM	16
Σχήμα 4 Διεπαφή WES Registry	36
Σχήμα 5 Διεπαφή Canvas Registry	42

Περίληψη

Το παρόν παραδοτέο (Π12.2) περιγράφει τον σχεδιασμό και την ανάπτυξη ενός ευφυούς μηχανισμού δημιουργίας μητρώων ασθενών στο πλαίσιο της Ενότητας Εργασίας 12 του έργου. Ο μηχανισμός αυτός λειτουργεί σε πλήρη διαλειτουργικότητα με τον Έξυπνο Ηλεκτρονικό Φάκελο Υγείας (Π12.1) και ενσωματώνει δυνατότητες για αξιολόγηση τεχνολογιών υγείας (Health Technology Assessment - HTA).

Τα μητρώα ασθενών αποτελούν θεμελιώδη εργαλεία για την έρευνα, τη χάραξη πολιτικών υγείας και τη βελτίωση της κλινικής πρακτικής. Στην εποχή της Ιατρικής Ακριβείας, η ανάγκη για μητρώα που συνδυάζουν κλινικά και γενετικά δεδομένα γίνεται επιτακτική. Το σύστημα που αναπτύχθηκε απαντά σε αυτή την ανάγκη, παρέχοντας μια ευέλικτη πλατφόρμα για τη δημιουργία, διαχείριση και ανάλυση μητρώων με γενετικά δεδομένα.

Η αρχιτεκτονική του συστήματος βασίζεται σε μια δυο επίπεδα: το επιχειρησιακό επίπεδο χρησιμοποιεί HL7 FHIR R4 για την κλινική λειτουργία, ενώ το αναλυτικό επίπεδο αξιοποιεί το OMOP Common Data Model με την επέκταση Genomic CDM για έρευνα και αναλύσεις. Αυτή η προσέγγιση επιτρέπει τόσο την πρωτογενή χρήση δεδομένων στην κλινική πράξη όσο και τη δευτερογενή χρήση για έρευνα και αξιολόγηση τεχνολογιών υγείας.

Στα πλαίσια του έργου υλοποιήθηκαν δύο πιλοτικά μητρώα: Το ένα περιλαμβάνει περίπου 300 ασθενείς με σπάνια νοσήματα (κυρίως νευρολογικά και καρδιολογικά) οι οποίοι παρέμεναν αδιάγνωστοι και αναλύθηκαν με Whole Exome Sequencing με στόχο τον εντοπισμό του υπεύθυνου γονιδίου και τη θέση μοριακής διάγνωσης. Για κάθε ασθενή καταγράφονται δημογραφικά στοιχεία, φαινοτυπικά χαρακτηριστικά με ηλικία εμφάνισης συμπτωμάτων, γενετικά ευρήματα (γονίδιο, παραλλαγή, ταξινόμηση ACMG, ζυγωτία), αποτέλεσμα διαγνωστικής ανάλυσης (positive/inconclusive/negative). Το **CanVaS Registry** (Cancer Variation reSource) περιλαμβάνει γενετικά δεδομένα από 7.363 Έλληνες με ατομικό ή οικογενειακό ιστορικό κακοήθειας, με περίπου 24.000 σπάνιες παραλλαγές σε 97 γονίδια προδιάθεσης στον καρκίνο.

1. Εισαγωγή: Μητρώα Ασθενών στην Ιατρική Ακριβείας

1.1 Ορισμός και Έννοια των Μητρώων Ασθενών

Τα μητρώα ασθενών αποτελούν οργανωμένα συστήματα συλλογής, αποθήκευσης και ανάλυσης δεδομένων υγείας για πληθυσμούς που προσδιορίζονται από συγκεκριμένη νόσο, κατάσταση ή έκθεση. Σε αντίθεση με τους ηλεκτρονικούς φακέλους υγείας που εστιάζουν στην παροχή φροντίδας σε ατομικό επίπεδο, τα μητρώα συγκεντρώνουν δεδομένα από πολλούς ασθενείς με κοινά χαρακτηριστικά, επιτρέποντας την εξαγωγή συμπερασμάτων σε επίπεδο πληθυσμού.

Η έννοια του μητρώου ασθενών έχει εξελιχθεί σημαντικά τις τελευταίες δεκαετίες. Από απλές βάσεις δεδομένων με δημογραφικά στοιχεία και διαγνώσεις, τα σύγχρονα μητρώα έχουν μετατραπεί σε σύνθετα πληροφοριακά συστήματα που ενσωματώνουν κλινικά δεδομένα, βιοδείκτες, απεικονιστικές εξετάσεις, γενετικές πληροφορίες και δεδομένα αποτελεσμάτων. Αυτή η εξέλιξη αντικατοπτρίζει τη μετάβαση από την περιγραφική επιδημιολογία στην αναλυτική έρευνα και την παραγωγή Real World Evidence (RWE).

Τα μητρώα διακρίνονται σε διάφορες κατηγορίες ανάλογα με τον σκοπό τους. Τα μητρώα νοσημάτων (disease registries) παρακολουθούν ασθενείς με συγκεκριμένη πάθηση, όπως καρκίνο ή σπάνια νοσήματα. Τα μητρώα διαδικασιών (procedure registries) καταγράφουν εκβάσεις συγκεκριμένων ιατρικών παρεμβάσεων. Τα πληθυσμιακά μητρώα (population registries) καλύπτουν ολόκληρους πληθυσμούς για σκοπούς δημόσιας υγείας. Τέλος, τα μητρώα προϊόντων (product registries) παρακολουθούν την ασφάλεια και αποτελεσματικότητα φαρμάκων ή ιατροτεχνολογικών προϊόντων μετά την κυκλοφορία τους.

1.2 Ο Ρόλος των Μητρώων στο Σύγχρονο Σύστημα Υγείας

1.2.1 Έρευνα και Επιδημιολογία

Τα μητρώα ασθενών αποτελούν ανεκτίμητη πηγή δεδομένων για την επιδημιολογική έρευνα. Επιτρέπουν τον υπολογισμό επίπτωσης και επιπολασμού νοσημάτων, την αναγνώριση παραγόντων κινδύνου, τη μελέτη της φυσικής ιστορίας νόσων και την αξιολόγηση της αποτελεσματικότητας θεραπειών σε πραγματικές συνθήκες. Σε αντίθεση με τις τυχαίοποιημένες κλινικές δοκιμές που διεξάγονται σε ελεγχόμενο περιβάλλον με αυστηρά κριτήρια επιλογής, τα μητρώα παρέχουν δεδομένα από την καθημερινή κλινική πρακτική, αντικατοπτρίζοντας την πραγματική ποικιλομορφία των ασθενών και των θεραπευτικών προσεγγίσεων.

Η μακροχρόνια παρακολούθηση που προσφέρουν τα μητρώα είναι ιδιαίτερα πολύτιμη για τη μελέτη χρόνιων νοσημάτων και σπάνιων εκβάσεων. Ενώ οι κλινικές δοκιμές συνήθως περιορίζονται σε λίγα έτη παρακολούθησης, τα μητρώα μπορούν να συλλέγουν δεδομένα για δεκαετίες, αποκαλύπτοντας μακροπρόθεσμες τάσεις και όψιμες επιπλοκές.

1.2.2 Analytics και Real World Evidence

Η ανάλυση δεδομένων μητρώων παράγει RWE, δηλαδή τεκμήρια για τη χρήση και τα πιθανά οφέλη ή κινδύνους ιατρικών προϊόντων που προέρχονται από πηγές εκτός παραδοσιακών κλινικών δοκιμών. Η RWE έχει αποκτήσει κεντρική θέση στη λήψη αποφάσεων στον τομέα της υγείας, καθώς οι ρυθμιστικές αρχές όπως ο FDA και ο EMA αναγνωρίζουν όλο και περισσότερο την αξία της.

Τα σύγχρονα εργαλεία ανάλυσης δεδομένων, συμπεριλαμβανομένων των τεχνικών μηχανικής μάθησης, επιτρέπουν την εξαγωγή γνώσης από τα μητρώα με τρόπους που δεν ήταν εφικτοί στο

παρελθόν. Η αναγνώριση μοτίβων, η πρόβλεψη εκβάσεων και η στρωματοποίηση κινδύνου αποτελούν εφαρμογές που μετασχηματίζουν τα δεδομένα μητρώων σε δράσιμη κλινική πληροφορία.

1.2.3 Πολιτικές Υγείας και Health Technology Assessment

Τα μητρώα διαδραματίζουν κρίσιμο ρόλο στη χάραξη πολιτικών υγείας και στην αξιολόγηση τεχνολογιών υγείας (ΗΤΑ). Οι φορείς ΗΤΑ, όπως το NICE στο Ηνωμένο Βασίλειο και το IQWiG στη Γερμανία, χρησιμοποιούν δεδομένα μητρώων για να αξιολογήσουν την κλινική αποτελεσματικότητα και τη σχέση κόστους-αποτελεσματικότητας νέων θεραπειών.

Η Αξιολόγηση Τεχνολογιών Υγείας έχει αυξανόμενη ανάγκη για δεδομένα πραγματικής κλινικής πρακτικής προκειμένου να εξετάσει την κλινική αποτελεσματικότητα πέρα από το ελεγχόμενο περιβάλλον των κλινικών δοκιμών και να πραγματοποιήσει οικονομική αξιολόγηση βασισμένη σε πραγματικά κόστη και αποτελέσματα. Τα μητρώα μπορούν να προσθέσουν στις κλασικές μεθόδους ΗΤΑ πραγματικά δεδομένα ασφάλειας, πληροφορίες για σπάνια γεγονότα, μακροχρόνια δεδομένα παρακολούθησης και ακριβή περιγραφή του πληθυσμού-στόχου.

Επιπλέον, τα μητρώα υποστηρίζουν τον σχεδιασμό υπηρεσιών υγείας μέσω της παροχής δεδομένων για τον επιπολασμό νοσημάτων, τη χρήση υπηρεσιών και τις ανάγκες σε πόρους. Αυτές οι πληροφορίες είναι απαραίτητες για τον ορθολογικό σχεδιασμό και την κατανομή πόρων σε εθνικό και περιφερειακό επίπεδο.

1.3 Παραδείγματα Μητρώων Ασθενών

1.3.1 Μητρώα στην Ελλάδα

Η Ελλάδα διαθέτει αρκετά σημαντικά μητρώα ασθενών, αν και το τοπίο παραμένει κατακερματισμένο. Το Εθνικό Μητρώο Νεοπλασιών¹ που τηρείται από το Υπουργείο Υγείας και την ΗΔΙΚΑ Α.Ε., καταγράφει όλες τις περιπτώσεις καρκίνου στη χώρα και αποτελεί βασικό εργαλείο για την παρακολούθηση της επιδημιολογίας του καρκίνου.

Το Μητρώο Ασθενών Σακχαρώδη Διαβήτη² του ΕΟΠΥΥ καταγράφει ασθενείς με διαβήτη τύπου 1 και 2 πανελλαδικά, με περισσότερους από 424.000 εγγεγραμμένους ασθενείς. Επιπλέον, το Εθνικό Μητρώο Αιμοδοτών³ που διαχειρίζεται η ΕΔΥΤΕ Α.Ε. σε συνεργασία με το Εθνικό Κέντρο Αιμοδοσίας, αποτελεί ενιαίο κεντρικό σύστημα καταγραφής εθελοντών αιμοδοτών με κάλυψη του 95% των συλλόγων εθελοντών αιμοδοτών.

1.3.2 Διεθνή Παραδείγματα

Διεθνώς, υπάρχουν πολλά επιτυχημένα παραδείγματα μητρώων που έχουν συμβάλει σημαντικά στην ιατρική γνώση. Το SEER⁴ (Surveillance, Epidemiology, and End Results) στις ΗΠΑ αποτελεί ένα από τα πιο ολοκληρωμένα μητρώα καρκίνου παγκοσμίως, καλύπτοντας περίπου το 45,9% του αμερικανικού πληθυσμού και παρέχοντας δεδομένα για επίπτωση, επιβίωση και τάσεις θνησιμότητας.

Το Orphanet⁵ αποτελεί την ευρωπαϊκή βάση δεδομένων αναφοράς για σπάνια νοσήματα, ενώ τα European Reference Networks⁶ (ERNs) συντονίζουν μητρώα για συγκεκριμένες κατηγορίες σπάνιων

¹ <https://www.cancer-registry.gov.gr/>

² <https://eservices.eopyy.gov.gr>

³ <https://www.blooddonorregistry.gr/>

⁴ <https://seer.cancer.gov/>

⁵ <https://www.orpha.net/>

⁶ https://health.ec.europa.eu/rare-diseases-and-european-reference-networks/european-reference-networks_en

παθήσεων. Το UK Biobank⁷, με περισσότερους από 500.000 συμμετέχοντες και πλούσια γενετικά δεδομένα, αποτελεί πρότυπο για ασθενών.

Οι σκανδιναβικές χώρες, ιδιαίτερα η Δανία και η Σουηδία, διαθέτουν ολοκληρωμένα εθνικά μητρώα που καλύπτουν σχεδόν το σύνολο του πληθυσμού και διασυνδέονται μέσω μοναδικών αριθμών ταυτοποίησης. Αυτά τα μητρώα έχουν αποδειχθεί ανεκτίμητα για την επιδημιολογική έρευνα και τη μελέτη μακροχρόνιων εκβάσεων.

1.4 Η Ανάγκη για Μητρώα με Γενετικά Δεδομένα

1.4.1 Περιορισμοί των Παραδοσιακών Μητρώων

Τα παραδοσιακά μητρώα ασθενών, αν και πολύτιμα, παρουσιάζουν σημαντικούς περιορισμούς στην εποχή της Ιατρικής Ακριβείας. Η κατηγοριοποίηση ασθενών βασίζεται κυρίως σε κλινικά κριτήρια και διαγνώσεις, αγνοώντας τη μοριακή ετερογένεια που υποκρύπτεται πίσω από κοινές κλινικές εκδηλώσεις. Δύο ασθενείς με την ίδια κλινική διάγνωση μπορεί να έχουν εντελώς διαφορετικό μοριακό υπόστρωμα και, κατά συνέπεια, διαφορετική πρόγνωση και ανταπόκριση σε θεραπείες.

Η έλλειψη γενετικών δεδομένων περιορίζει επίσης τη δυνατότητα αναγνώρισης υποομάδων ασθενών που θα ωφεληθούν από συγκεκριμένες θεραπείες, τη μελέτη φαρμακογενετικών αλληλεπιδράσεων και την κατανόηση της κληρονομικότητας νοσημάτων σε οικογένειες.

1.4.2 Μητρώα στην Εποχή της Ιατρικής Ακριβείας

Η Ιατρική Ακριβείας απαιτεί τη σύνδεση γονοτύπου και φαινοτύπου σε κλίμακα πληθυσμού. Μητρώα που ενσωματώνουν γενετικά δεδομένα επιτρέπουν την κατανόηση του γενετικού υποβάθρου νοσημάτων στον ελληνικό πληθυσμό, την αναγνώριση πληθυσμιακά-ειδικών παραλλαγών (founder mutations) με κλινική σημασία, τη μελέτη της συσχέτισης γονοτύπου με κλινικά χαρακτηριστικά, πρόγνωση και ανταπόκριση σε θεραπείες, καθώς και την ανάπτυξη μοντέλων πρόβλεψης κινδύνου προσαρμοσμένων στα χαρακτηριστικά του ελληνικού πληθυσμού.

Η ανάγκη για πληθυσμιακά-ειδικά δεδομένα είναι ιδιαίτερα σημαντική, καθώς οι περισσότερες γενετικές μελέτες έχουν διεξαχθεί σε πληθυσμούς βορειοευρωπαϊκής καταγωγής. Ο ελληνικός πληθυσμός, λόγω της ιστορικής του απομόνωσης σε ορισμένες περιοχές, παρουσιάζει ιδιαίτερα γενετικά χαρακτηριστικά και founder mutations που δεν έχουν μελετηθεί επαρκώς.

1.4.3 Στόχοι του Παρόντος Παραδοτέου

Το παραδοτέο Π12.2 απαντά στην ανάγκη για σύγχρονα μητρώα που ενσωματώνουν γενετικά δεδομένα, παρέχοντας έναν ευφυή μηχανισμό για τη δημιουργία και διαχείριση τέτοιων μητρώων. Ο μηχανισμός αυτός σχεδιάστηκε ώστε να λειτουργεί σε πλήρη διαλειτουργικότητα με τον Έξυπνο Ηλεκτρονικό Φάκελο Υγείας (Π12.1), αξιοποιώντας τα δεδομένα που συλλέγονται κατά την κλινική πρακτική για την αυτόματη ή ημιαυτόματη τροφοδότηση μητρώων.

⁷ <https://www.ukbiobank.ac.uk/>

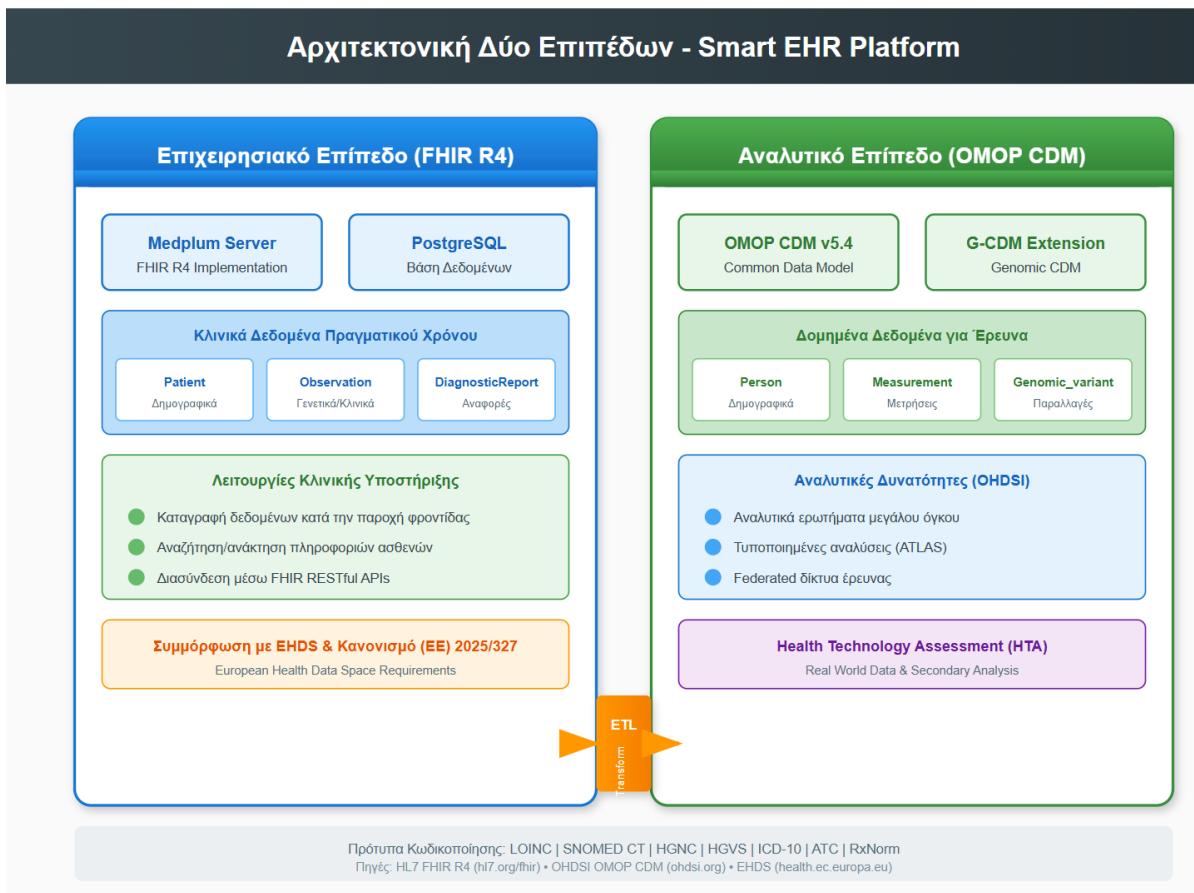
2. Αρχιτεκτονική Συστήματος Μητρώων

2.1 Σχεδιαστικές Αρχές

Η αρχιτεκτονική του συστήματος μητρώων σχεδιάστηκε με γνώμονα την κάλυψη δύο διακριτών αλλά αλληλένδετων αναγκών: την υποστήριξη της κλινικής λειτουργίας και την παροχή δυνατοτήτων για έρευνα και ανάλυση. Αυτές οι δύο ανάγκες έχουν διαφορετικές απαιτήσεις σε επίπεδο μοντέλου δεδομένων, ερωτημάτων και απόδοσης, γεγονός που οδήγησε στην υιοθέτηση μιας αρχιτεκτονικής δύο επιπέδων (dual-layer architecture), η οποία απεικονίζεται στο Σχήμα 1.

Κεντρική αρχή του σχεδιασμού αποτελεί η διαλειτουργικότητα. Το σύστημα βασίζεται σε ανοιχτά πρότυπα τόσο για την κλινική πλευρά (HL7 FHIR R4) όσο και για την αναλυτική (OMOP CDM), διασφαλίζοντας τη δυνατότητα ανταλλαγής δεδομένων με άλλα συστήματα και τη συμμετοχή σε διεθνή δίκτυα έρευνας.

2.2 Αρχιτεκτονική Δύο Επιπέδων



Σχήμα 1 Αρχιτεκτονική Smart EHR Platform

2.2.1 Επιχειρησιακό Επίπεδο (FHIR R4)

Το επιχειρησιακό επίπεδο χρησιμοποιεί το πρότυπο HL7 FHIR R4 ⁸ για την αποθήκευση και διαχείριση κλινικών δεδομένων σε πραγματικό χρόνο. Αυτό το επίπεδο υποστηρίζει την καθημερινή κλινική λειτουργία, επιτρέποντας την καταγραφή δεδομένων κατά την παροχή φροντίδας, την αναζήτηση και

⁸ <https://www.hl7.org/fhir/>

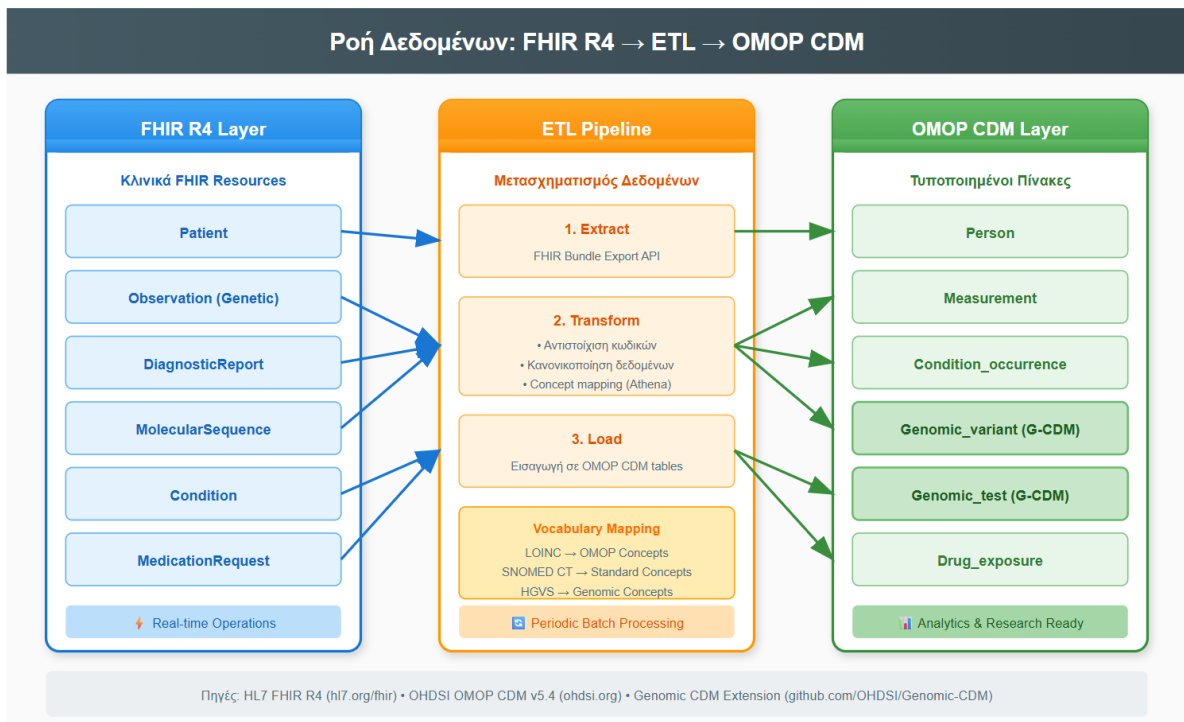
ανάκτηση πληροφοριών για μεμονωμένους ασθενείς, τη διασύνδεση με άλλα κλινικά συστήματα μέσω FHIR APIs και την υποστήριξη κλινικής απόφασης σε πραγματικό χρόνο.

Η επιλογή του FHIR R4 βασίστηκε στην ευρεία υιοθέτησή του ως πρότυπο διαλειτουργικότητας, τη συμβατότητα με τις απαιτήσεις του European Health Data Space ⁹(EHDS) και την ύπαρξη ώριμων υλοποιήσεων και εργαλείων.

2.2.2 Αναλυτικό Επίπεδο (OMOP CDM)

Το αναλυτικό επίπεδο χρησιμοποιεί το OMOP Common Data Model (CDM) για τη δομημένη αποθήκευση δεδομένων με σκοπό την έρευνα και την ανάλυση. Το OMOP CDM, που αναπτύχθηκε από τον οργανισμό OHDSI (Observational Health Data Sciences and Informatics), έχει σχεδιαστεί για να υποστηρίζει αναλυτικά ερωτήματα σε μεγάλους όγκους δεδομένων, τη σύγκριση μεταξύ διαφορετικών πηγών δεδομένων, την εκτέλεση τυποποιημένων αναλύσεων και τη συμμετοχή σε federated δίκτυα έρευνας.

Η μετάβαση από το FHIR στο OMOP, όπως παρουσιάζεται στο Σχήμα 2, επιτυγχάνεται μέσω διαδικασιών ETL (Extract-Transform-Load) που εκτελούνται περιοδικά, μετασχηματίζοντας τα κλινικά δεδομένα σε μορφή κατάλληλη για ανάλυση.



Σχήμα 2 Ροή μετάβασης δεδομένων από το FHIR στο OMOP

2.2.3 Γιατί Χρειάζονται και τα Δύο Επίπεδα

Η απόφαση για διατήρηση δύο ξεχωριστών επιπέδων, αντί για μια ενιαία βάση δεδομένων, βασίζεται σε θεμελιώδεις διαφορές στις απαιτήσεις. Το κλινικό περιβάλλον απαιτεί γρήγορη πρόσβαση σε δεδομένα μεμονωμένων ασθενών, υποστήριξη για συχνές ενημερώσεις και transactional integrity.

⁹ https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space-regulation-ehds_en

Αντίθετα, το αναλυτικό περιβάλλον απαιτεί τη δυνατότητα εκτέλεσης σύνθετων ερωτημάτων σε εκατομμύρια εγγραφές, τυποποιημένη δομή για cross-institutional συγκρίσεις και βελτιστοποίηση για αναλυτικά workloads.

Επιπλέον, ο διαχωρισμός επιτρέπει την εφαρμογή διαφορετικών πολιτικών ασφάλειας και πρόσβασης. Τα δεδομένα στο αναλυτικό επίπεδο μπορούν να υποστούν ψευδωνυμοποίηση ή ανωνυμοποίηση, διευκολύνοντας τη χρήση τους για έρευνα.

2.3 Δεδομένα που Καταγράφονται

2.3.1 Κλινικά Δεδομένα

Το σύστημα καταγράφει ένα ευρύ φάσμα κλινικών δεδομένων που είναι απαραίτητα για τη λειτουργία των μητρώων. Τα δημογραφικά στοιχεία περιλαμβάνουν ηλικία, φύλο, γεωγραφική περιοχή και εθνικότητα, τα οποία είναι απαραίτητα για επιδημιολογικές αναλύσεις και τη μελέτη γεωγραφικής κατανομής νοσημάτων.

Οι διαγνώσεις κωδικοποιούνται με ICD-10 και SNOMED CT, ενώ για σπάνια γενετικά νοσήματα χρησιμοποιείται επιπλέον το OMIM¹⁰. Τα κλινικά συμπτώματα περιγράφονται με όρους HPO¹¹ (Human Phenotype Ontology), επιτρέποντας το λεπτομερή φαινοτυπικό χαρακτηρισμό.

Καταγράφονται επίσης οι θεραπείες και παρεμβάσεις και τα αποτελέσματα εργαστηριακών εξετάσεων. Κάθε καταχώρηση συνοδεύεται από πληροφορίες προέλευσης (data provenance), συμπεριλαμβανομένης της πηγής, της ημερομηνίας καταχώρησης και της έκδοσης.

2.3.2 Γενετικά Δεδομένα

Τα γενετικά δεδομένα αποτελούν κεντρικό στοιχείο των μητρώων που δημιουργούνται μέσω του παρόντος μηχανισμού. Καταγράφονται οι γενετικές παραλλαγές που έχουν εντοπιστεί, με πλήρη περιγραφή σε μορφή HGVS¹² (Human Genome Variation Society) τόσο σε επίπεδο DNA όσο και πρωτεΐνης. Κάθε παραλλαγή συνοδεύεται από την ταξινόμησή της κατά ACMG (Pathogenic, Likely Pathogenic, VUS, Likely Benign, Benign) και τα στοιχεία τεκμηρίωσης της ταξινόμησης.

Επιπλέον, καταγράφονται φαρμακογενετικά δεδομένα για τα βασικά γονίδια που επηρεάζουν τον μεταβολισμό φαρμάκων, καθώς και Polygenic Risk Scores (PRS) για κοινά πολυπαραγοντικά νοσήματα. Για τις παραλλαγές επίσης καταγράφεται επιπλέον ο τύπος διαμονής των ασθενών.

2.4 OMOP CDM και Genomic CDM Extension

2.4.1 Βασικοί Πίνακες OMOP

Το OMOP Common Data Model παρέχει μια τυποποιημένη δομή για την αναπαράσταση κλινικών δεδομένων. Οι βασικοί πίνακες που χρησιμοποιούνται περιλαμβάνουν τον PERSON για δημογραφικά στοιχεία, τον CONDITION_OCCURRENCE για διαγνώσεις, τον DRUG_EXPOSURE για φαρμακευτικές αγωγές, τον MEASUREMENT για εργαστηριακά αποτελέσματα και άλλες μετρήσεις, τον OBSERVATION για κλινικές παρατηρήσεις και τον PROCEDURE_OCCURRENCE για ιατρικές διαδικασίες.

¹⁰ <https://www.omim.org/>

¹¹ <https://hpo.jax.org/>

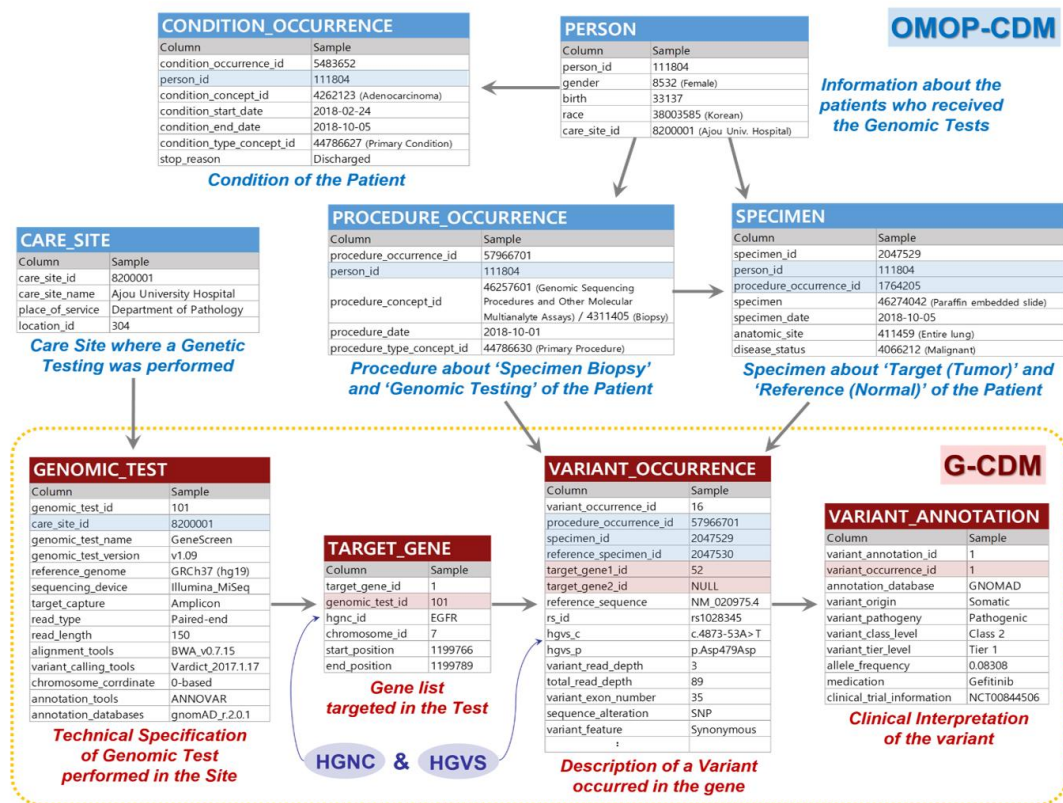
¹² <https://hgvs-nomenclature.org/>

Όλα τα δεδομένα κωδικοποιούνται με χρήση τυποποιημένων vocabularies του OMOP, επιτρέποντας τη σύγκριση και ανάλυση μεταξύ διαφορετικών πηγών δεδομένων.

2.4.2 Genomic CDM Extension

Για την αναπαράσταση γενετικών δεδομένων, το σύστημα υιοθετεί την επέκταση Genomic CDM¹³ που αναπτύχθηκε από την κοινότητα OHDSI. Η επέκταση αυτή προσθέτει εξειδικευμένους πίνακες για γενετικά δεδομένα, διατηρώντας τη συμβατότητα με το βασικό OMOP CDM.

Ο πίνακας TARGET_GENE περιέχει πληροφορίες για τα γονίδια που εξετάστηκαν σε κάθε ασθενή, συμπεριλαμβανομένου του τύπου εξέτασης και του αποτελέσματος (θετικό, αρνητικό, αδιευκρίνιστο). Ο πίνακας VARIANT_OCCURRENCE καταγράφει τις συγκεκριμένες παραλλαγές που εντοπίστηκαν, με αναφορά στη θέση, το reference και alternate allele. Ο πίνακας VARIANT_ANNOTATION περιέχει τις ερμηνείες και ταξινομήσεις των παραλλαγών.



Σχήμα 3 Αποθετήριο για την ανάπτυξη του genomic module - CDM

2.4.3 Mapping FHIR προς OMOP

Η μετατροπή δεδομένων από το FHIR στο OMOP αποτελεί κρίσιμη διαδικασία για τη λειτουργία του αναλυτικού επιπέδου. Ο μετασχηματισμός ακολουθεί καθορισμένους κανόνες αντιστοίχισης για κάθε τύπο δεδομένων.

Οι FHIR Patient resources μετατρέπονται σε εγγραφές στον πίνακα PERSON, με αντιστοίχιση των δημογραφικών πεδίων. Οι Condition resources γίνονται CONDITION_OCCURRENCE, με μετατροπή

¹³ <https://github.com/OHDSI/Genomic-CDM>

των κωδικών ICD-10 στα αντίστοιχα OMOP concepts. Οι Observation resources για γενετικά ευρήματα διασπώνται στους εξειδικευμένους πίνακες του Genomic CDM.

Πίνακας 1 FHIR σε OMOP mapping

FHIR Resource	OMOP Table	Genomic CDM Table
Patient	PERSON	-
Condition	CONDITION_OCCURRENCE	-
Observation (Gene)	MEASUREMENT	TARGET_GENE
Observation (Variant)	MEASUREMENT	VARIANT_OCCURRENCE
DiagnosticReport	PROCEDURE_OCCURRENCE	-
RiskAssessment (PRS)	MEASUREMENT	MEASUREMENT (extended)

2.5 Διαλειτουργικότητα

2.5.1 FHIR APIs για Ανταλλαγή Δεδομένων

Το σύστημα παρέχει πλήρη FHIR R4 API για την ανταλλαγή δεδομένων με εξωτερικά συστήματα. Αυτό επιτρέπει τη διασύνδεση με ηλεκτρονικούς φακέλους υγείας, εργαστηριακά πληροφοριακά συστήματα και της πηγές κλινικών δεδομένων. Η υιοθέτηση του FHIR διασφαλίζει τη συμβατότητα με της απαιτήσεις του European Health Data Space.

2.5.2 OMOP για Cross-Institutional Analytics

Η χρήση του OMOP CDM επιτρέπει τη συμμετοχή σε διεθνή δίκτυα έρευνας της το OHDSI. Μέσω του OMOP, είναι δυνατή η εκτέλεση κοινών αναλύσεων σε δεδομένα από διαφορετικά ιδρύματα χωρίς να απαιτείται μεταφορά των δεδομένων (federated analytics).

2.5.3 Συμμόρφωση με EHDS

Η αρχιτεκτονική έχει σχεδιαστεί ώστε να υποστηρίζει τις απαιτήσεις του European Health Data Space για πρωτογενή και δευτερογενή χρήση δεδομένων υγείας. Αυτό περιλαμβάνει τη χρήση τυποποιημένων μορφών ανταλλαγής, την υποστήριξη για διασυννοριακή πρόσβαση και τη συμμόρφωση με τα ευρωπαϊκά πλαίσια διακυβέρνησης δεδομένων.

3. Κύκλος Ζωής Μητρώου

Ο κύκλος ζωής της μητρώου ασθενών αποτελεί μια δυναμική διαδικασία που ξεκινά από τον αρχικό σχεδιασμό και εκτείνεται μέχρι τη συνεχή λειτουργία και εξέλιξη του. Το παρόν κεφάλαιο περιγράφει αναλυτικά κάθε φάση αυτού του κύκλου, της μηχανισμούς εμπλουτισμού δεδομένων, της διαδικασίες διασφάλισης ποιότητας, καθώς και της τρόπους παρακολούθησης και εξέλιξης του μητρώου διαχρονικά. Η κατανόηση αυτού του κύκλου είναι θεμελιώδης για την επιτυχή υλοποίηση και βιωσιμότητα οποιουδήποτε μητρώου ασθενών, ιδιαίτερα στο πλαίσιο της Ιατρικής Ακρίβειας όπου η ποιότητα και η πληρότητα των γενετικών δεδομένων είναι κρίσιμες.

3.1 Ορισμός Μητρώου (Registry Definition Phase)

Η φάση ορισμού του μητρώου αποτελεί το θεμέλιο για την επιτυχημένη λειτουργία του συστήματος. Κατά τη διάρκεια της φάσης, καθορίζονται οι βασικές παράμετροι που θα καθοδηγήσουν τη συλλογή, αποθήκευση και ανάλυση δεδομένων. Οι αποφάσεις που λαμβάνονται σε αυτό το στάδιο έχουν μακροπρόθεσμες επιπτώσεις στην ποιότητα και χρησιμότητα του μητρώου, γι' αυτό απαιτείται ιδιαίτερη προσοχή και συστηματική προσέγγιση.

3.1.1 Ορισμός Κοόρτης (Cohort Definition)

Ο ορισμός της κοόρτης αποτελεί την πρώτη και κρίσιμη απόφαση στη δημιουργία της μητρώου. Η κοόρτη καθορίζει με ακρίβεια ποιοι ασθενείς θα συμπεριληφθούν στο μητρώο και υπό ποιες προϋποθέσεις. Της σαφής και αναπαραγωγίμος ορισμός κοόρτης διασφαλίζει τη συνέπεια της συλλογής δεδομένων και επιτρέπει τη σύγκριση μεταξύ διαφορετικών χρονικών περιόδων ή ιδρυμάτων.

3.1.1.1 Κριτήρια Εισαγωγής (Inclusion Criteria)

Τα κριτήρια εισαγωγής καθορίζουν τις απαραίτητες προϋποθέσεις για την ένταξη ενός ασθενούς στο μητρώο. Το σύστημα υποστηρίζει σύνθετους ορισμούς κριτηρίων που μπορούν να συνδυάσουν:

- Κλινικές διαγνώσεις: Κωδικοποιημένες με ICD-10, SNOMED CT, ή OMIM για γενετικά νοσήματα. Παράδειγμα: «Ασθενείς με διάγνωση ICD-10 G71.0 (Μυϊκή δυστροφία) ή OMIM 310200 (Duchenne muscular dystrophy)»
- Φαινοτυπικά χαρακτηριστικά: Καταγεγραμμένα με όρους Human Phenotype Ontology (HPO). Παράδειγμα: «Ασθενείς με HP:0001263 (Global developmental delay) ΚΑΙ HP:0001250 (Seizures)»
- Γενετικά κριτήρια: Παρουσία παραλλαγών σε συγκεκριμένα γονίδια ή με συγκεκριμένη ταξινόμηση ACMG. Παράδειγμα: «Ασθενείς με Pathogenic ή Likely Pathogenic παραλλαγή σε γονίδια BRCA1 ή BRCA2»
- Δημογραφικά κριτήρια: Ηλικία, φύλο, γεωγραφική περιοχή καταγωγής. Παράδειγμα: «Ενήλικες (≥18 ετών) γυναίκες με καταγωγή από Κρήτη»
- Χρονικά κριτήρια: Διάστημα διάγνωσης, ημερομηνία εξέτασης. Παράδειγμα: «Ασθενείς με γενετική εξέταση μετά την 1η Ιανουαρίου 2020»

3.1.1.2 Κριτήρια Αποκλεισμού (Exclusion Criteria)

Τα κριτήρια αποκλεισμού αφαιρούν από την κοόρτη ασθενείς που πληρούν τα κριτήρια εισαγωγής αλλά δεν πρέπει να συμπεριληφθούν για συγκεκριμένους λόγους:

- Ποιότητα δεδομένων: Ασθενείς με ελλιπή κρίσιμα δεδομένα (π.χ. απουσία επιβεβαιωμένης διάγνωσης, ελλιπής γενετική ανάλυση)
- Κλινικές αντενδείξεις: Συννοσηρότητες που επηρεάζουν την ερμηνεία αποτελεσμάτων
- Συγκατάθεση: Ασθενείς που έχουν αρνηθεί τη συμμετοχή ή έχουν ανακαλέσει τη συγκατάθεσή τους
- Διπλότυπες εγγραφές: Ασθενείς που έχουν ήδη καταχωρηθεί με διαφορετικό αναγνωριστικό

3.1.1.3 Γλώσσα Ορισμού Κοόρτης βάσει FHIR

Η γλώσσα ορισμού κοόρτης του συστήματος βασίζεται σε FHIR Search Parameters και υποστηρίζει σύνθετες λογικές εκφράσεις. Κάθε κριτήριο μεταφράζεται σε FHIR query που μπορεί να εκτελεστεί στον FHIR Server για τον προσδιορισμό των επιλέξιμων ασθενών. Η χρήση τυποποιημένων FHIR queries διασφαλίζει την αναπαραγωγιμότητα και τη δυνατότητα μεταφοράς ορισμών κοόρτης μεταξύ συστημάτων, Πίνακας 2. Επιπλέον, το σύστημα υποστηρίζει ATLAS Cohort Definitions για συμβατότητα με το οικοσύστημα OHDSI.

Πίνακας 2 Παραδείγματα Ορισμών Κοόρτης

Τύπος Μητρώου	Κριτήρια Εισαγωγής	FHIR Query Pattern
WES Registry	Ασθενείς με WES ανάλυση και νευρολογικά ή καρδιολογικά συμπτώματα	Observation?code=http://loinc.org 81247-9&_has:DiagnosticReport:result:status=final
CanVaS Registry	Άτομα με γενετική εξέταση για καρκίνο και BRCA1/BRCA2 ανάλυση	Observation?code=http://loinc.org 48018-6&value-concept=HGNC:1100,HGNC:1101
Pharmacogenomics	Ασθενείς με φαρμακογενετική ανάλυση για CYP2D6, CYP2C19	Observation?code=http://loinc.org 81247-9&component-code=http://loinc.org 48018-6

3.1.2 Προδιαγραφές Dataset (Dataset Specifications)

Μετά τον ορισμό της κοόρτης, καθορίζονται τα πεδία δεδομένων που θα συλλέγονται για κάθε ασθενή. Οι προδιαγραφές dataset αποτελούν ένα δομημένο σχήμα που περιγράφει κάθε μεταβλητή, τον τύπο της, τα επιτρεπόμενα value sets και τους κανόνες validation, Πίνακας 3. Η σαφής τεκμηρίωση των προδιαγραφών είναι απαραίτητη για τη συνέπεια της συλλογής δεδομένων και τη διαλειτουργικότητα με άλλα συστήματα.

Κατηγοριοποίηση Πεδίων Δεδομένων

Τα πεδία δεδομένων κατηγοριοποιούνται βάσει της κρισιμότητάς τους:

Πίνακας 3 Κρισιμότητα και Κατηγοριοποίηση

Κατηγορία	Χαρακτηριστικό	Παραδείγματα
Υποχρεωτικά (Required)	Πρέπει να υπάρχουν για κάθε εγγραφή	Patient ID, Gender, Diagnosis, Gene tested
Συνιστώμενα (Recommended)	Ιδανικά θα πρέπει να συμπληρώνονται	Age at onset, HPO terms, Geographic origin
Προαιρετικά (Optional)	Καταγράφονται όταν είναι διαθέσιμα	Family history details, Treatment history
Υπολογιζόμενα (Derived)	Υπολογίζονται αυτόματα από άλλα πεδία	Age at diagnosis, Diagnostic yield metrics

Τυποποιημένα Value Sets και Συστήματα Κωδικοποίησης

Για τη διασφάλιση της διαλειτουργικότητας και της ποιότητας δεδομένων, όλα τα κωδικοποιημένα πεδία χρησιμοποιούν διεθνώς αναγνωρισμένα συστήματα κωδικοποίησης, όπως παρουσιάζονται στον Πίνακα 4:

Πίνακας 4 Διαλειτουργικότητα και Συστήματα Κωδικοποίησης

Σύστημα Κωδικοποίησης	URI	Χρήση στο Μητρώο
LOINC	http://loinc.org	Εργαστηριακές εξετάσεις, γενετικά components
SNOMED CT	http://snomed.info/sct	Κλινικά ευρήματα, διαγνώσεις
ICD-10	http://hl7.org/fhir/sid/icd-10	Νοσήματα, κατηγοριοποίηση
HPO	http://purl.obolibrary.org/obo/hp.owl	Φαινοτυπικά χαρακτηριστικά
HGNC	http://www.genenames.org	Ονοματολογία γονιδίων
OMIM	http://www.omim.org	Γενετικά νοσήματα
ClinVar	http://www.ncbi.nlm.nih.gov/clinvar	Ταξινόμηση παραλλαγών

3.1.3 Δείκτες Ποιότητας και Αποτελεσμάτων (Quality and Outcome Indicators)

Κάθε μητρώο συνοδεύεται από ένα σύνολο Key Performance Indicators (KPIs) που επιτρέπουν την παρακολούθηση τόσο της ποιότητας των δεδομένων όσο και των κλινικών αποτελεσμάτων, Πίνακας 5. Οι δείκτες αυτοί υπολογίζονται αυτόματα και παρουσιάζονται σε dashboards για τους διαχειριστές του μητρώου και τους κλινικούς.

Δείκτες Ποιότητας Δεδομένων

Πίνακας 5 Παρακολούθηση ποιότητας Δεδομένων

Δείκτης	Περιγραφή	Στόχος
Completeness Score	Ποσοστό συμπλήρωσης υποχρεωτικών πεδίων	≥95%
Timeliness	Μέσος χρόνος από κλινική επαφή έως καταχώρηση	<7 ημέρες
Validity Rate	Ποσοστό εγγραφών που περνούν validation	≥98%
Consistency Score	Ποσοστό εγγραφών χωρίς λογικές αντιφάσεις	≥99%
Coding Accuracy	Ποσοστό σωστής χρήσης κωδικολογίων	≥97%
Duplicate Rate	Ποσοστό διπλότυπων εγγραφών	<0.5%

Κλινικοί Δείκτες Αποτελεσμάτων

Οι κλινικοί δείκτες, Πίνακας 6, εξαρτώνται από τον τύπο του μητρώου και παρακολουθούν τα αποτελέσματα της διαγνωστικής και θεραπευτικής διαδικασίας:

Πίνακας 6 Δείκτες και αποτελέσματα

Δείκτης	Τύπος Μητρώου	Τρόπος Υπολογισμού
Diagnostic Yield	Σπάνια νοσήματα	$(P + LP \text{ results}) / \text{Total tests} \times 100\%$
VUS Resolution Rate	Γενετικά μητρώα	$\text{VUS reclassified} / \text{Total VUS} \times 100\%$
Time to Diagnosis	Σπάνια νοσήματα	Diagnosis date - First symptom date

Founder Mutation Rate	Πληθυσμιακά μητρώα	$\text{Founder P/LP} / \text{Total P/LP} \times 100\%$
Actionable Findings	Κλινικά μητρώα	$\text{Clinically actionable} / \text{Total findings} \times 100\%$

3.1.4 Διακυβέρνηση Μητρώου (Registry Governance)

Το πλαίσιο διακυβέρνησης καθορίζει τους ρόλους, τις ευθύνες και τις διαδικασίες για τη διαχείριση του μητρώου. Μια σαφής δομή διακυβέρνησης είναι απαραίτητη για τη βιωσιμότητα του μητρώου και την εξασφάλιση της εμπιστοσύνης όλων των εμπλεκομένων μερών, Πίνακας 7.

Πίνακας 7 Ρόλοι Διακυβέρνησης Μητρώου

Ρόλος	Ευθύνες	Συχνότητα Δράσης
Registry Owner	Στρατηγική κατεύθυνση, τελική έγκριση αλλαγών, διασύνδεση με stakeholders	Μηνιαία
Scientific Committee	Ορισμός κλινικών κριτηρίων, έγκριση τροποποιήσεων, επιστημονική εποπτεία	Τριμηνιαία
Data Manager	Καθημερινή διαχείριση, έλεγχος ποιότητας, επίλυση προβλημάτων δεδομένων	Καθημερινή
Data Entry Staff	Καταχώρηση δεδομένων, επικοινωνία με κλινικούς για διευκρινίσεις	Καθημερινή
IT Administrator	Τεχνική υποστήριξη, backups, ασφάλεια συστήματος	Καθημερινή
DPO / Privacy Officer	Συμμόρφωση GDPR, διαχείριση αιτημάτων υποκειμένων	Κατά απαίτηση

3.2 Μηχανισμοί Εμπλουτισμού Δεδομένων (Data Enrichment Mechanisms)

Η αξία ενός μητρώου εξαρτάται άμεσα από τη συνεχή τροφοδότησή του με ποιοτικά δεδομένα. Ο μηχανισμός που αναπτύχθηκε υποστηρίζει τρεις διαφορετικές προσεγγίσεις για τον εμπλουτισμό δεδομένων, οι οποίες μπορούν να συνδυαστούν ανάλογα με τις ανάγκες και τις τεχνικές δυνατότητες κάθε περίπτωσης. Η επιλογή της κατάλληλης μεθόδου εξαρτάται από παράγοντες όπως η δυνατότητα διασύνδεσης συστημάτων, η συχνότητα ενημερώσεων, οι απαιτήσεις ασφάλειας και οι διαθέσιμοι πόροι.

3.2.1 Event-Driven (Live) Enrichment

Ο εμπλουτισμός σε πραγματικό χρόνο αποτελεί την πλέον προηγμένη μέθοδο τροφοδοσίας μητρώων, επιτρέποντας την άμεση ενημέρωση χωρίς χρονικές καθυστερήσεις. Η υλοποίηση βασίζεται στον μηχανισμό FHIR Subscriptions, ο οποίος επιτρέπει σε ένα σύστημα να «εγγραφεί» για ειδοποιήσεις όταν συμβαίνουν συγκεκριμένα events στον FHIR Server.

Μηχανισμός FHIR Subscriptions

Το FHIR R4 Subscriptions framework παρέχει έναν τυποποιημένο τρόπο για real-time ειδοποιήσεις. Ο μηχανισμός λειτουργεί ως εξής:

- **Ορισμός Subscription:** Δημιουργείται ένα Subscription resource που καθορίζει τα κριτήρια ενεργοποίησης (π.χ. «Observation?code=http://loinc.org|53037-8» για νέες ACMG ταξινομήσεις) και τον τρόπο ειδοποίησης (REST hook, WebSocket, ή email).
- **Event Detection:** Όταν δημιουργείται, ενημερώνεται ή διαγράφεται ένα resource που ταιριάζει με τα κριτήρια, ο FHIR Server ενεργοποιεί την ειδοποίηση.

- **Notification Delivery:** Η ειδοποίηση αποστέλλεται στο καθορισμένο endpoint, περιλαμβάνοντας το resource ή αναφορά σε αυτό.
- **Processing:** Το Registry Engine επεξεργάζεται την ειδοποίηση, ελέγχει αν ο ασθενής πληροί τα κριτήρια του μητρώου, και ενημερώνει τα δεδομένα, Πίνακας 8.

Πίνακας 8 Περιπτώσεις Χρήσης Event-Driven Enrichment

Event Type	Trigger Condition	Registry Action
Νέα γενετική εξέταση	DiagnosticReport.status = final	Έλεγχος eligibility, εισαγωγή σε μητρώο
Νέα παραλλαγή	Observation (variant) created	Ενημέρωση γενετικού προφίλ ασθενούς
Αλλαγή ταξινόμησης	Observation.component (ACMG) updated	Re-evaluation αποτελέσματος
Νέα διάγνωση	Condition.verificationStatus = confirmed	Ενημέρωση φαινοτύπου
Ανάκληση συγκατάθεσης	Consent.status = rejected	Αφαίρεση από μητρώο

3.2.2 Batch ETL (Extract-Transform-Load)

Η περιοδική εισαγωγή δεδομένων μέσω διαδικασιών ETL αποτελεί την πλέον διαδεδομένη μέθοδο εμπλουτισμού μητρώων, ιδιαίτερα όταν η πηγή δεδομένων δεν υποστηρίζει real-time διασύνδεση. Η προσέγγιση αυτή είναι κατάλληλη για legacy συστήματα, εξωτερικές πηγές δεδομένων (CSV, Excel), αρχεία από εργαστήρια, και ιστορικά δεδομένα που χρειάζεται να μεταφερθούν στο μητρώο.

Φάσεις Διαδικασίας ETL

Extract (Εξαγωγή): Κατά τη φάση εξαγωγής, τα δεδομένα ανακτώνται από την πηγή τους. Το σύστημα υποστηρίζει ποικιλία μορφών εισόδου, συμπεριλαμβανομένων CSV, Excel (xlsx), JSON, XML, και LOVD export format. Ειδικό parsers έχουν αναπτυχθεί για τις συνηθέστερες μορφές γενετικών δεδομένων, όπως VCF files για raw variant data και LOVD exports για δεδομένα από βάσεις γενετικών παραλλαγών.

Transform (Μετασηματισμός): Η φάση μετασηματισμού περιλαμβάνει την κανονικοποίηση και τυποποίηση των δεδομένων. Οι κρίσιμες λειτουργίες περιλαμβάνουν: αντιστοίχιση ελεύθερου κειμένου σε τυποποιημένους κωδικούς (π.χ. HPO resolution για συμπτώματα), μετατροπή HGVS notation σε τυποποιημένη μορφή, επικύρωση γενετικών παραλλαγών με εξωτερικές βάσεις (ClinVar, gnomAD), υπολογισμό derived fields, και εφαρμογή business rules για τον καθορισμό του overall interpretation.

Load (Φόρτωση): Η τελική φάση περιλαμβάνει τη δημιουργία FHIR Bundles και την εισαγωγή τους στον FHIR Server. Η διαδικασία διαχειρίζεται τα dependencies μεταξύ resources (π.χ. Patient πρέπει να δημιουργηθεί πριν τα Observations), χειρίζεται duplicates και updates, και καταγράφει λεπτομερή logs για auditing.

Ο προγραμματισμός των ETL Jobs παρουσιάζεται στον Πίνακα 9.

Πίνακας 9 Προγραμματισμός ETL Jobs

Συχνότητα	Cron Expression	Περιπτώσεις Χρήσης
Καθημερινή	0 2 * * *	Ενημερώσεις από EHR, εργαστηριακά αποτελέσματα

Εβδομαδιαία	0 3 * * 0	Εξωτερικές πηγές, report generation
Μηνιαία	0 4 1 * * *	Historical data reconciliation, ClinVar updates
On-demand	Manual trigger	Initial load, bulk imports, emergency updates

3.2.3 Federated Approach (Ομοσπονδιακή Προσέγγιση)

Η federated προσέγγιση αντιπροσωπεύει ένα καινοτόμο μοντέλο διαχείρισης δεδομένων που επιτρέπει αναλύσεις χωρίς κεντρική συλλογή πρωτογενών δεδομένων, Πίνακας 10. Αυτό είναι ιδιαίτερα σημαντικό στο πλαίσιο του GDPR και των αυστηρών απαιτήσεων προστασίας γενετικών δεδομένων, καθώς και για τη συνεργασία μεταξύ ιδρυμάτων που δεν επιθυμούν ή δεν επιτρέπεται να μοιραστούν τα δεδομένα τους κεντρικά.

Αρχιτεκτονική Federated Network

Στο federated μοντέλο, κάθε συμμετέχον ιδρυμα (node) διατηρεί τα δεδομένα του τοπικά σε τυποποιημένη μορφή OMOP CDM. Η αρχιτεκτονική περιλαμβάνει τα ακόλουθα συστατικά:

- Local Data Node: Κάθε ιδρυμα διατηρεί τη δική του OMOP database με τα δεδομένα του. Η βάση παραμένει πλήρως υπό τον έλεγχο του ιδρύματος.
- Central Coordinator: Ένα κεντρικό σημείο συντονισμού (χωρίς πρόσβαση σε πρωτογενή δεδομένα) διαχειρίζεται τα scripts ανάλυσης και συγκεντρώνει τα αποτελέσματα.
- Analysis Scripts: Τυποποιημένα R ή SQL scripts από τη βιβλιοθήκη OHDSI (ATLAS, HADES) που εκτελούνται ομοιόμορφα σε όλους τους κόμβους.
- Aggregate Results: Μόνο συγκεντρωτικά αποτελέσματα (counts, statistics, coefficients) επιστρέφονται στον coordinator για meta-analysis.

Πίνακας 10 Σύγκριση Κεντρικής και Federated Προσέγγισης

Χαρακτηριστικό	Κεντρική Συλλογή	Federated
Τοποθεσία δεδομένων	Κεντρική βάση	Τοπικά σε κάθε ιδρυμα
Έλεγχος δεδομένων	Κεντρικός	Διατηρείται τοπικά
GDPR συμμόρφωση	Απαιτεί νομική βάση	Ευκολότερη συμμόρφωση
Πολυπλοκότητα setup	Χαμηλότερη	Υψηλότερη
Ευελιξία αναλύσεων	Πλήρης	Περιορισμένη σε προορισμένα scripts
Κλιμάκωση	Περιορίζεται από storage	Σχεδόν απεριόριστη

3.3 Αυτόματος έναντι Χειροκίνητου Εμπλουτισμού

Η επιλογή μεταξύ αυτόματου και χειροκίνητου εμπλουτισμού εξαρτάται από πολλούς παράγοντες, συμπεριλαμβανομένης της διαθεσιμότητας τεχνικών πόρων, της ποιότητας των πηγών δεδομένων, των απαιτήσεων ακρίβειας, και του κόστους υλοποίησης. Στην πράξη, τα περισσότερα μητρώα χρησιμοποιούν έναν συνδυασμό των δύο προσεγγίσεων.

3.3.1 Αυτόματος Εμπλουτισμός

Ο αυτόματος εμπλουτισμός είναι εφικτός όταν υπάρχει πλήρης διαλειτουργικότητα μεταξύ των συστημάτων-πηγών και του μητρώου. Προϋποθέτει δομημένα δεδομένα στην πηγή, σαφείς κανόνες μετασχηματισμού, και αξιόπιστη υποδομή δικτύου. Τα FHIR-native συστήματα, όπως ο Έξυπνος ΗΦΥ που αναπτύχθηκε στο Π12.1, υποστηρίζουν πλήρως αυτό το μοντέλο, επιτρέποντας τη ροή δεδομένων χωρίς ανθρώπινη παρέμβαση.

Πλεονεκτήματα Αυτόματου Εμπλουτισμού

- Ταχύτητα: Τα δεδομένα είναι διαθέσιμα στο μητρώο σχεδόν αμέσως μετά την καταχώρησή τους στην πηγή
- Συνέπεια: Οι κανόνες μετασχηματισμού εφαρμόζονται ομοιόμορφα σε όλες τις εγγραφές
- Κλιμακωσιμότητα: Μπορεί να χειριστεί μεγάλους όγκους δεδομένων χωρίς αύξηση προσωπικού
- Μείωση σφαλμάτων: Εξαλείφονται τα ανθρώπινα λάθη κατά την καταχώρηση

3.3.2 Χειροκίνητος Εμπλουτισμός

Όταν η αυτόματη διασύνδεση δεν είναι δυνατή, η χειροκίνητη καταχώρηση παραμένει μια έγκυρη επιλογή, ιδιαίτερα για μικρά μητρώα ή για δεδομένα που απαιτούν ανθρώπινη κρίση. Το σύστημα παρέχει φιλικές προς τον χρήστη διεπαφές για χειροκίνητη καταχώρηση, με ενσωματωμένους ελέγχους validation και autocomplete για τυποποιημένα πεδία.

Υβριδική Προσέγγιση

Στην πράξη, η βέλτιστη προσέγγιση είναι συχνά υβριδική. Τα δομημένα δεδομένα (δημογραφικά, κωδικοποιημένες διαγνώσεις, εργαστηριακά αποτελέσματα) μπορούν να εισάγονται αυτόματα, ενώ τα πολύπλοκα δεδομένα που απαιτούν κλινική ερμηνεία (π.χ. σύνθετη περιγραφή φαινοτύπου, αξιολόγηση VUS) μπορούν να απαιτούν ανθρώπινη επαλήθευση, Πίνακας 11.

Πίνακας 11 Σύγκριση Μεθόδων Εμπλουτισμού

Παράμετρος	Αυτόματος	Χειροκίνητος
Ταχύτητα	Άμεση ή scheduled	Εξαρτάται από διαθέσιμο προσωπικό
Κόστος setup	Υψηλό (αρχικά)	Χαμηλό
Κόστος λειτουργίας	Χαμηλό	Υψηλό (συνεχές)
Κλιμακωσιμότητα	Εξαιρετική	Περιορισμένη
Ευελιξία	Περιορισμένη από κανόνες	Υψηλή
Κατάλληλο για	Δομημένα, μεγάλο όγκου	Σύνθετα, απαιτούν κρίση

3.4 Διασφάλιση Ποιότητας Δεδομένων (Data Quality Assurance)

Η ποιότητα των δεδομένων αποτελεί κρίσιμο παράγοντα για την αξιοπιστία και χρησιμότητα κάθε μητρώου. Ένα μητρώο με δεδομένα χαμηλής ποιότητας μπορεί να οδηγήσει σε εσφαλμένα συμπεράσματα και κακές κλινικές αποφάσεις. Το σύστημα υλοποιεί πολυεπίπεδους μηχανισμούς διασφάλισης ποιότητας που καλύπτουν ολόκληρο τον κύκλο ζωής των δεδομένων.

3.4.1 Κανόνες Επικύρωσης (Validation Rules)

Κατά την εισαγωγή δεδομένων, εφαρμόζεται ένα πλήρες σύνολο κανόνων επικύρωσης που διασφαλίζει την ακεραιότητα και συνέπεια των δεδομένων, Πίνακας 12:

Πίνακας 12 Κανόνες και κατηγορίες ελέγχου

Κατηγορία Ελέγχου	Περιγραφή	Παράδειγμα
Format Validation	Έλεγχος ότι τα δεδομένα ακολουθούν το αναμενόμενο format	HGVS notation: c.[0-9]+[A-Z]>[A-Z]

Value Set Validation	Έλεγχος ότι οι τιμές ανήκουν σε επιτρεπόμενα value sets	Gender ∈ {male, female, unknown, other}
Range Validation	Έλεγχος ότι οι αριθμητικές τιμές είναι εντός λογικών ορίων	Age at onset: 0-120 years
Referential Integrity	Έλεγχος ότι τα references δείχνουν σε υπαρκτά resources	Patient reference exists
Cross-field Validation	Έλεγχος λογικής συνέπειας μεταξύ πεδίων	Diagnosis date ≥ Birth date
Uniqueness Validation	Έλεγχος για αποφυγή διπλότυπων	Patient ID unique per registry

3.4.2 Παρακολούθηση Ποιότητας (Quality Monitoring)

Πέρα από τους ελέγχους κατά την εισαγωγή, το σύστημα εκτελεί περιοδικές αναλύσεις ποιότητας που εντοπίζουν προβλήματα που μπορεί να μην είναι εμφανή σε επίπεδο μεμονωμένης εγγραφής αλλά γίνονται ορατά σε επίπεδο συνόλου δεδομένων. Αυτές περιλαμβάνουν ελέγχους πληρότητας (completeness checks) που υπολογίζουν τα ποσοστά συμπλήρωσης για κάθε πεδίο και εντοπίζουν τάσεις, ελέγχους συνέπειας (consistency checks) που αναζητούν λογικές αντιφάσεις στα δεδομένα, και ανάλυση κατανομής (distribution analysis) που εντοπίζει ακραίες τιμές (outliers) και ασυνήθιστες κατανομές.

3.5 Εξέλιξη Μητρώου και Διαχείριση Εκδόσεων

Τα μητρώα δεν είναι στατικά συστήματα αλλά εξελίσσονται διαρκώς για να ανταποκριθούν σε νέες επιστημονικές γνώσεις, κανονιστικές απαιτήσεις, και κλινικές ανάγκες. Το σύστημα υποστηρίζει τη διαχείριση αυτής της εξέλιξης μέσω αυστηρών διαδικασιών versioning και change management.

3.5.1 Εξέλιξη Σχήματος Δεδομένων

Η εξέλιξη του σχήματος δεδομένων (schema evolution) απαιτεί προσεκτική διαχείριση για να διασφαλιστεί η συνέπεια με τα υπάρχοντα δεδομένα. Το σύστημα υποστηρίζει τις ακόλουθες τύπους αλλαγών: προσθήκη νέων πεδίων (backward compatible), αφαίρεση πεδίων (με περίοδο deprecation), αλλαγή τύπου δεδομένων (με μετατροπή), ενημέρωση value sets (με mapping παλαιών τιμών), και αλλαγή κανόνων validation.

3.5.2 Διαδικασία Διαχείρισης Αλλαγών

Κάθε αλλαγή στον ορισμό του μητρώου ακολουθεί μια τυποποιημένη διαδικασία που διασφαλίζει την ποιότητα και τη συνέπεια:

- Αίτημα αλλαγής: Τεκμηρίωση της αλλαγής, της αιτιολόγησής της, και της επίπτωσης στα υπάρχοντα δεδομένα
- Αξιολόγηση: Ανάλυση της επίπτωσης στα υπάρχοντα δεδομένα και τις τρέχουσες αναλύσεις
- Έγκριση: Έγκριση από το Scientific Committee ή τον Registry Owner
- Υλοποίηση: Εφαρμογή της αλλαγής με migration scripts αν απαιτείται
- Τεκμηρίωση: Ενημέρωση του data dictionary και του changelog
- Επικοινωνία: Ενημέρωση των χρηστών για τις αλλαγές

Η συστηματική διαχείριση του κύκλου ζωής του μητρώου, από τον αρχικό ορισμό έως τη συνεχή λειτουργία και εξέλιξη, αποτελεί τη βάση για τη δημιουργία αξιόπιστων και χρήσιμων μητρώων ασθενών. Οι μηχανισμοί που περιγράφηκαν σε αυτό το κεφάλαιο διασφαλίζουν την ποιότητα, συνέπεια και βιωσιμότητα των μητρώων που δημιουργούνται μέσω του συστήματος Π12.2.

4. Ασφάλεια και Συμμόρφωση

4.1 Συμμόρφωση με GDPR

Τα μητρώα ασθενών περιέχουν ευαίσθητα προσωπικά δεδομένα που υπόκεινται στις αυστηρές απαιτήσεις του Γενικού Κανονισμού Προστασίας Δεδομένων (GDPR). Το σύστημα σχεδιάστηκε με την προστασία της ιδιωτικότητας ως θεμελιώδη αρχή (privacy by design).

4.1.1 Νομική Βάση Επεξεργασίας

Η επεξεργασία δεδομένων υγείας για σκοπούς μητρώων μπορεί να βασίζεται σε διαφορετικές νομικές βάσεις ανάλογα με τον σκοπό. Για την πρωτογενή χρήση στην παροχή φροντίδας, η επεξεργασία δικαιολογείται από το άρθρο 9(2)(h) του GDPR για σκοπούς ιατρικής διάγνωσης και θεραπείας. Για τη δευτερογενή χρήση σε έρευνα, απαιτείται είτε ρητή συγκατάθεση είτε εφαρμογή του άρθρου 9(2)(j) για σκοπούς δημόσιου συμφέροντος στον τομέα της δημόσιας υγείας.

Το σύστημα διατηρεί πλήρες αρχείο της νομικής βάσης για κάθε επεξεργασία και υποστηρίζει τη διαχείριση συγκαταθέσεων όπου αυτή απαιτείται.

4.1.2 Ελαχιστοποίηση Δεδομένων

Η αρχή της ελαχιστοποίησης δεδομένων εφαρμόζεται σε όλα τα επίπεδα του συστήματος. Συλλέγονται μόνο τα δεδομένα που είναι απολύτως απαραίτητα για τον σκοπό κάθε μητρώου. Οι ορισμοί μητρώων περιλαμβάνουν ρητή τεκμηρίωση για την αναγκαιότητα κάθε πεδίου.

Στο αναλυτικό επίπεδο (OMOP), εφαρμόζονται τεχνικές ψευδωνυμοποίησης και de-identification, αφαιρώντας ή μετασχηματίζοντας τα άμεσα αναγνωριστικά στοιχεία.

4.1.3 Δικαιώματα Υποκειμένων

Το σύστημα υποστηρίζει την άσκηση των δικαιωμάτων που παρέχει ο GDPR στα υποκείμενα των δεδομένων. Το δικαίωμα πρόσβασης υλοποιείται μέσω δυνατότητας εξαγωγής όλων των δεδομένων που αφορούν έναν ασθενή σε αναγνώσιμη μορφή. Το δικαίωμα διαγραφής υποστηρίζεται με μηχανισμούς που διασφαλίζουν την πλήρη αφαίρεση δεδομένων από όλα τα επίπεδα του συστήματος, με τις εξαιρέσεις που προβλέπει ο νόμος για δεδομένα υγείας.

4.2 Έλεγχος Πρόσβασης

4.2.1 Role-Based Access Control (RBAC)

Το σύστημα υλοποιεί ένα πολυεπίπεδο μοντέλο ελέγχου πρόσβασης βασισμένο σε ρόλους. Κάθε χρήστης αντιστοιχίζεται σε έναν ή περισσότερους ρόλους, και κάθε ρόλος έχει καθορισμένα δικαιώματα πρόσβασης, Πίνακας 13.

Πίνακας 13 Ρόλοι και Πρόσβαση

Ρόλος	Πρόσβαση	Λειτουργίες
Κλινικός	Δικοί του ασθενείς	Ανάγνωση, καταχώρηση
Ερευνητής	De-identified δεδομένα	Αναλύσεις, εξαγωγές
Registry Manager	Πλήρης στο μητρώο	Διαχείριση, QC, αναφορές
Administrator	Πλήρης στο σύστημα	Διαχείριση χρηστών, audit

4.2.2 Διαχείριση Συγκαταθέσεων

Για μητρώα που απαιτούν συγκατάθεση ασθενών, το σύστημα παρέχει πλήρη διαχείριση του κύκλου ζωής της συγκατάθεσης. Καταγράφεται η ημερομηνία και ο τρόπος λήψης συγκατάθεσης, το πεδίο

εφαρμογής της (ποια δεδομένα, για ποιους σκοπούς), τυχόν περιορισμοί ή εξαιρέσεις, και η κατάσταση (ενεργή, ανακληθείσα, ληγμένη).

Η ανάκληση συγκατάθεσης ενεργοποιεί αυτόματα τις απαραίτητες διαδικασίες για τη διακοπή επεξεργασίας ή τη διαγραφή δεδομένων, ανάλογα με τις απαιτήσεις.

4.3 Audit και Ιχνηλασιμότητα

4.3.1 Audit Trail

Το σύστημα διατηρεί πλήρες audit trail για όλες τις ενέργειες που αφορούν δεδομένα ασθενών. Κάθε καταχώρηση στο audit log περιλαμβάνει τον χρήστη που εκτέλεσε την ενέργεια, τον τύπο ενέργειας (δημιουργία, ανάγνωση, ενημέρωση, διαγραφή), το αντικείμενο της ενέργειας (resource type και ID), την ακριβή χρονοσήμανση, καθώς και την IP διεύθυνση και άλλα στοιχεία συνεδρίας.

Τα audit logs είναι αμετάβλητα (immutable) και διατηρούνται για το χρονικό διάστημα που απαιτεί η νομοθεσία.

4.3.2 Data Provenance

Πέρα από το audit trail, το σύστημα καταγράφει την προέλευση κάθε δεδομένου (data provenance). Αυτό περιλαμβάνει την πηγή του δεδομένου (χρήστης, σύστημα, ETL διαδικασία), τη μέθοδο εισαγωγής (χειροκίνητη, αυτόματη, batch), την έκδοση σε περίπτωση ενημερώσεων, και τις σχέσεις με άλλα δεδομένα.

Η πλήρης ιχνηλασιμότητα είναι απαραίτητη για τη διασφάλιση της ποιότητας δεδομένων και την υποστήριξη ερευνητικών αναλύσεων που απαιτούν γνώση της προέλευσης των δεδομένων.

5. Τεχνική Υλοποίηση

Το παρόν κεφάλαιο περιγράφει την τεχνική υλοποίηση του μηχανισμού δημιουργίας μητρώων ασθενών στα πλαίσια του έργου TAEDR-0539180. Η αρχιτεκτονική βασίζεται σε μια προσέγγιση ETL (Extract-Transform-Load) με έμφαση στη διαλειτουργικότητα μέσω FHIR R4 resources, επιτρέποντας την ενοποίηση δεδομένων από πολλαπλές πηγές με διατήρηση της προέλευσης και δυνατότητα διάκρισης μεταξύ διαφορετικών μητρώων.

5.1 Registry Engine

Ο Registry Engine αποτελεί τον πυρήνα του μηχανισμού δημιουργίας μητρώων. Στην υλοποίηση που αναπτύχθηκε, ο Engine είναι υπεύθυνος για τη διαχείριση του κύκλου ζωής των δεδομένων από την εισαγωγή έως την αποθήκευση στο FHIR repository.

5.1.1 Σύστημα Επισήμανσης Μητρώων (Registry Tagging)

Κεντρικό στοιχείο της υλοποίησης αποτελεί το σύστημα επισήμανσης (tagging) που επιτρέπει τη διάκριση δεδομένων ανάλογα με το μητρώο προέλευσης. Κατά την εισαγωγή δεδομένων, κάθε FHIR resource επισημαίνεται με μοναδικό tag που προσδιορίζει το μητρώο από το οποίο προέρχεται. Αυτή η προσέγγιση επιτρέπει την αποθήκευση δεδομένων από πολλαπλά μητρώα στον ίδιο FHIR Server, διατηρώντας παράλληλα πλήρη διαχωρισμό και δυνατότητα απομόνωσης κατά την αναζήτηση.

Για κάθε μητρώο ορίζεται ένα μοναδικό Tag System URI και αντίστοιχοι κωδικοί για κάθε τύπο resource:

Πίνακας 14 Μητρώο και Tagging

Μητρώο	Tag System URI	Tags
WES Registry	http://wes-database.example.org/tags	wes-registry, wes-variant, wes-gene, wes-report
CanVaS Registry	http://canvas.rrp.demokritos.gr/tags	canvas-registry, canvas-variant, canvas-gene, canvas-report

Η χρήση του μηχανισμού meta.tag του FHIR R4 διασφαλίζει τη συμβατότητα με το πρότυπο και επιτρέπει αποδοτικές αναζητήσεις μέσω του _tag search parameter. Για παράδειγμα, η ανάκτηση όλων των ασθενών του WES Registry πραγματοποιείται με: GET /Patient?_tag=http://wes-database.example.org/tags|wes-registry

5.1.2 Αρχιτεκτονική ETL Pipeline

Η εισαγωγή δεδομένων στα μητρώα υλοποιείται μέσω εξειδικευμένων ETL (Extract-Transform-Load) pipelines που μετασχηματίζουν δεδομένα από ετερογενείς πηγές σε τυποποιημένους FHIR R4 resources. Κάθε pipeline ακολουθεί μια τριμερή αρχιτεκτονική:

Extract Phase: Ανάγνωση δεδομένων από τις πηγές. Υποστηρίζονται μορφές Excel (xlsx), CSV, JSON, και εξειδικευμένες μορφές όπως LOVD exports για γενετικές παραλλαγές.

Transform Phase: Μετασχηματισμός και κανονικοποίηση. Περιλαμβάνει αντιστοίχιση όρων σε τυποποιημένα κωδικολόγια (LOINC, SNOMED CT, HPO, HGNC), επικύρωση HGVS notation, και υπολογισμό derived fields.

Load Phase: Δημιουργία FHIR Transaction Bundles και εισαγωγή στον FHIR Server. Τα bundles δομούνται με σωστή σειρά για ικανοποίηση των dependencies μεταξύ resources.

Στην τρέχουσα υλοποίηση έχουν αναπτυχθεί πλήρεις ETL pipelines για τα δύο πιλοτικά μητρώα του έργου. Ο WES Pipeline επεξεργάζεται δεδομένα από αρχεία Excel με γενετικά δεδομένα ασθενών σπάνιων νοσημάτων, ενώ ο CanVaS Pipeline εισάγει δεδομένα από τη βάση LOVD του ΕΚΕΦΕ

«Δημόκριτος». Και οι δύο pipelines ενσωματώνουν κλήσεις σε εξωτερικά APIs όπως το JAX Ontology API για την επίλυση HPO όρων, με μηχανισμούς caching για βελτιστοποίηση της απόδοσης.

5.1.3 Cohort Builder

Η αρχιτεκτονική προβλέπει ένα Cohort Builder component για τον ορισμό κριτηρίων εισαγωγής και αποκλεισμού ασθενών σε μητρώα. Ο μηχανισμός υποστηρίζει τον ορισμό κριτηρίων βασισμένων σε διαγνώσεις (ICD-10, OMIM), κλινικά χαρακτηριστικά (HPO terms), γενετικά ευρήματα (γονίδια, ACMG classifications), και δημογραφικά στοιχεία. Οι ορισμοί κοορτών μεταφράζονται σε FHIR Search queries που μπορούν να εκτελεστούν στον FHIR Server, ενώ παράλληλα σχεδιάζεται η υποστήριξη ATLAS Cohort Definitions για συμβατότητα με το οικοσύστημα OHDSI.

5.1.4 Scheduler και Batch Processing

Ο Scheduler διαχειρίζεται τις προγραμματισμένες εργασίες του συστήματος, συμπεριλαμβανομένων των batch ETL jobs για εισαγωγή δεδομένων, των quality checks, και της παραγωγής στατιστικών αναφορών. Κάθε εργασία μπορεί να προγραμματιστεί με ευέλικτους κανόνες και συνοδεύεται από μηχανισμούς logging για παρακολούθηση της εκτέλεσης.

5.2 Registry Store

Το Registry Store αποτελεί την υποδομή αποθήκευσης δεδομένων, υλοποιώντας μια αρχιτεκτονική δύο επιπέδων (dual-layer architecture) που εξυπηρετεί διαφορετικές ανάγκες: το επιχειρησιακό επίπεδο για κλινική λειτουργία και το αναλυτικό επίπεδο για έρευνα.

5.2.1 FHIR Server (Operational Layer)

Το επιχειρησιακό επίπεδο βασίζεται στον Medplum FHIR Server, μια πλήρη υλοποίηση του προτύπου HL7 FHIR R4. Ο server παρέχει αποθήκευση σε PostgreSQL, RESTful API για CRUD operations, πλήρη υποστήριξη FHIR Search parameters, καθώς και μηχανισμούς authentication και authorization. Η επιλογή του Medplum βασίστηκε στην ωριμότητα της πλατφόρμας, την ευρεία υποστήριξη για γενωμικά δεδομένα, και τη συμβατότητα με τις απαιτήσεις του European Health Data Space.

Στην τρέχουσα υλοποίηση, ο FHIR Server περιέχει τα δεδομένα και των δύο πιλοτικών μητρώων:

Πίνακας 15 Μητρώο και Δεδομένα

Μητρώο	Αριθμός Ασθενών	FHIR Resources
WES Registry	~300 ασθενείς	Patient, Observation, DiagnosticReport, Condition, ServiceRequest, Specimen
CanVaS Registry	~ 7.400 άτομα	Patient, Observation (Variant, Gene), Condition, DiagnosticReport

Η διάκριση μεταξύ των μητρώων επιτυγχάνεται μέσω του συστήματος tagging που περιγράφηκε στην ενότητα 5.1.1, επιτρέποντας τόσο τη συνολική ανάλυση όσο και την απομονωμένη πρόσβαση σε δεδομένα συγκεκριμένου μητρώου.

5.2.2 OMOP Database (Analytics Layer)

Το αναλυτικό επίπεδο σχεδιάστηκε με βάση το OMOP Common Data Model (CDM) του OHDSI, συμπληρωμένο από την επέκταση Genomic CDM για γενετικά δεδομένα. Η αρχιτεκτονική αυτή επιτρέπει τη συμμετοχή σε federated δίκτυα έρευνας και τη διεξαγωγή τυποποιημένων αναλύσεων συμβατών με τα εργαλεία OHDSI (ATLAS, HADES).

Στο πλαίσιο του έργου έχει σχεδιαστεί η δομή της βάσης OMOP με τους βασικούς πίνακες (PERSON, CONDITION_OCCURRENCE, MEASUREMENT, PROCEDURE_OCCURRENCE) καθώς και τους εξειδικευμένους πίνακες του Genomic CDM (TARGET_GENE, VARIANT_OCCURRENCE, VARIANT_ANNOTATION). Η υλοποίηση proof-of-concept έχει ολοκληρωθεί και επιτρέπει τον μετασχηματισμό δεδομένων από FHIR σε OMOP format.

5.2.3 ETL Pipeline FHIR → OMOP

Η μεταφορά δεδομένων από το επιχειρησιακό επίπεδο (FHIR) στο αναλυτικό επίπεδο (OMOP) πραγματοποιείται μέσω διαδικασιών ETL. Η διαδικασία περιλαμβάνει την ανάκτηση FHIR resources, τον μετασχηματισμό σύμφωνα με κανόνες mapping, την αντιστοίχιση κωδικών (ICD-10, SNOMED, LOINC, HGNC) σε OMOP concept IDs, και τέλος τη φόρτωση στους αντίστοιχους πίνακες OMOP.

Πίνακας 16 Αντιστοίχιση FHIR Resources σε OMOP Tables

FHIR Resource	OMOP Table	Genomic CDM Table
Patient	PERSON, LOCATION	-
Condition	CONDITION_OCCURRENCE	-
Observation (Gene)	MEASUREMENT	TARGET_GENE
Observation (Variant)	MEASUREMENT	VARIANT_OCCURRENCE, VARIANT_ANNOTATION
DiagnosticReport	PROCEDURE_OCCURRENCE	-

5.3 Έλεγχοι Ποιότητας Δεδομένων

Η ποιότητα των δεδομένων είναι κρίσιμη για την αξιοπιστία των μητρώων. Το σύστημα υλοποιεί πολυεπίπεδους ελέγχους ποιότητας που εφαρμόζονται τόσο κατά την εισαγωγή δεδομένων όσο και περιοδικά.

5.3.1 Validation Rules

Κατά την εισαγωγή δεδομένων εφαρμόζονται κανόνες validation που περιλαμβάνουν έλεγχο τύπου δεδομένων και format (π.χ. ημερομηνίες, HGVS notation), έλεγχο επιτρεπόμενων τιμών και value sets σύμφωνα με διεθνή κωδικολογία, έλεγχο υποχρεωτικών πεδίων, καθώς και cross-field validation για λογική συνέπεια μεταξύ πεδίων. Στα ETL pipelines που αναπτύχθηκαν ενσωματώνονται οι παραπάνω έλεγχοι, με καταγραφή λεπτομερών logs για τις εγγραφές που αποτυγχάνουν στην επικύρωση.

5.3.2 Completeness Checks

Περιοδικά εκτελούνται έλεγχοι πληρότητας που υπολογίζουν το ποσοστό συμπλήρωσης για κάθε πεδίο, εντοπίζουν εγγραφές με ελλείπουσες κρίσιμες πληροφορίες, και παράγουν αναφορές για τους διαχειριστές μητρώων. Οι έλεγχοι αυτοί είναι ιδιαίτερα σημαντικοί για γενετικά δεδομένα όπου η απουσία πληροφοριών (π.χ. ACMG classification, HGVS notation) επηρεάζει άμεσα τη χρησιμότητα του μητρώου.

5.3.3 Consistency Checks

Οι έλεγχοι συνέπειας διασφαλίζουν ότι τα δεδομένα είναι λογικά συνεπή. Για παράδειγμα, ελέγχεται ότι η ημερομηνία διάγνωσης είναι μετά την ημερομηνία γέννησης, ότι οι κωδικοί διαγνώσεων είναι συνεπείς με τα κλινικά ευρήματα, και ότι οι γενετικές παραλλαγές αντιστοιχούν σε έγκυρες θέσεις στο γονιδίωμα. Επιπλέον, ο μηχανισμός tagging επιτρέπει τον έλεγχο ότι κάθε resource ανήκει σε ένα και μόνο μητρώο.

5.3.4 Reconciliation με Κλινικούς

Το σύστημα υποστηρίζει διαδικασίες επικύρωσης δεδομένων από κλινικούς ειδικούς. Οι εγγραφές που χρήζουν επανελέγχου επισημαίνονται και οι κλινικοί μπορούν να επιβεβαιώσουν ή να διορθώσουν τα δεδομένα. Αυτή η διαδικασία είναι ιδιαίτερα σημαντική για τα VUS (Variants of Uncertain Significance) όπου η ταξινόμηση μπορεί να αλλάξει με νέα επιστημονικά δεδομένα.

5.4 Dashboards και Αναφορές

5.4.1 Registry Overview Dashboard

Κάθε μητρώο συνοδεύεται από dashboard που παρουσιάζει τα βασικά στατιστικά. Το dashboard περιλαμβάνει τον συνολικό αριθμό εγγραφών και τις τάσεις εισαγωγών, τα δημογραφικά χαρακτηριστικά της κοόρτης (κατανομή ηλικίας, κατανομή φύλου), τους δείκτες ποιότητας δεδομένων, καθώς και τα κύρια κλινικά και γενετικά χαρακτηριστικά.

5.4.2 Clinical Indicators

Για κάθε μητρώο ορίζονται ειδικοί κλινικοί δείκτες που παρακολουθούνται μέσω του dashboard. Για το WES Registry, οι δείκτες περιλαμβάνουν το diagnostic yield (ποσοστό θετικών αποτελεσμάτων), το VUS rate, και την κατανομή ευρημάτων ανά γονίδιο. Για το CanVaS Registry, παρακολουθούνται η κατανομή παραλλαγών ανά ACMG classification, οι founder mutations, και οι φαινοτυπικές κατανομές.

5.4.3 Export Functionality

Το σύστημα παρέχει δυνατότητες εξαγωγής δεδομένων σε διάφορες μορφές, ανάλογα με τον σκοπό χρήσης. Υποστηρίζονται εξαγωγές σε CSV για απλές αναλύσεις, σε FHIR Bundle (JSON ή XML) για διαλειτουργικότητα με άλλα συστήματα, καθώς και σε μορφές συμβατές με εργαλεία OHDSI για συμμετοχή σε federated αναλύσεις. Όλες οι εξαγωγές υπόκεινται στους ελέγχους πρόσβασης και καταγράφονται στο audit trail.

5.5 Σύνοψη Κατάστασης Υλοποίησης

Ο παρακάτω πίνακας συνοψίζει την τρέχουσα κατάσταση υλοποίησης των τεχνικών συστατικών του μηχανισμού δημιουργίας μητρώων:

Πίνακας 17 Κατάσταση Υλοποίησης

Συστατικό	Κατάσταση	Σημειώσεις
FHIR Server (Medplum)	Υλοποιημένο	Πλήρης λειτουργία
Registry Tagging System	Υλοποιημένο	Χρήση σε WES & CanVaS
ETL Pipeline (WES)	Υλοποιημένο	~300 ασθενείς εισήχθησαν
ETL Pipeline (CanVaS)	Υλοποιημένο	7.363 άτομα εισήχθησαν
OMOP CDM Schema	Σχεδιασμένο	POC ολοκληρωμένο
Genomic CDM Extension	Σχεδιασμένο	POC ολοκληρωμένο
FHIR → OMOP ETL	Proof of Concept	Mapping ορισμένο
Cohort Builder	Αρχιτεκτονική	FHIR Search queries
Data Quality Checks	Υλοποιημένο	Ενσωμάτωση σε ETL

Η υλοποίηση επικεντρώθηκε στη δημιουργία μιας στιβαρής υποδομής για την εισαγωγή και διαχείριση δεδομένων μητρώων μέσω FHIR, με το αναλυτικό επίπεδο OMOP να σχεδιάζεται για μελλοντική επέκταση και συμμετοχή σε federated δίκτυα έρευνας. Η διαλειτουργικότητα μέσω FHIR R4 διασφαλίζει τη συμβατότητα με τις απαιτήσεις του European Health Data Space και επιτρέπει την εύκολη ενσωμάτωση νέων πηγών δεδομένων μέσω του μηχανισμού tagging.

6. Υλοποίηση στα Πλαίσια του Έργου

Στα πλαίσια του έργου υλοποιήθηκαν δύο πιλοτικά μητρώα που αποδεικνύουν τις δυνατότητες του μηχανισμού και παρέχουν πολύτιμα δεδομένα για την Ιατρική Ακριβείας στην Ελλάδα. Η υλοποίηση βασίζεται σε ένα κοινό μοντέλο δεδομένων FHIR R4 που επιτρέπει τη διαλειτουργικότητα μεταξύ των μητρώων, ενώ το αναλυτικό επίπεδο OMOP CDM με την επέκταση Genomic CDM υποστηρίζει τη δευτερογενή ανάλυση και τη συμμετοχή σε federated δίκτυα έρευνας.

6.1 Κοινό Μοντέλο Δεδομένων FHIR

Το κοινό μοντέλο δεδομένων αποτελεί τη βάση για την τυποποιημένη αναπαράσταση κλινικών και γενετικών δεδομένων σε όλα τα μητρώα του συστήματος. Η σχεδίαση ακολουθεί το πρότυπο HL7 FHIR R4 με επεκτάσεις ειδικά σχεδιασμένες για τις ανάγκες της Ιατρικής Ακριβείας και τα ιδιαίτερα χαρακτηριστικά του ελληνικού πληθυσμού.

6.1.1 Μοντελοποίηση Ασθενών (Patient Resource)

Κάθε ασθενής αναπαρίσταται ως FHIR Patient resource με τυποποιημένα πεδία για δημογραφικά στοιχεία και επεκτάσεις για πληροφορίες σχετικές με γενετικές αναλύσεις. Τα βασικά στοιχεία που καταγράφονται περιλαμβάνουν:

Πίνακας 18 Πεδία και Μοντελοποίηση

Πεδίο	FHIR Element	Περιγραφή
Αναγνωριστικό	Patient.identifier	Μοναδικός κωδικός μητρώου (WES ID ή CanVaS ID)
Φύλο	Patient.gender	male female unknown
Ημερομηνία γέννησης	Patient.birthDate	Για υπολογισμό ηλικίας έναρξης
Γεωγραφική καταγωγή	Patient.extension[geographic-origin]	Περιοχή καταγωγής γονέων
Συγγένεια γονέων	Patient.extension[consanguinity]	Boolean για αιμομιξία
Οικογενειακές σχέσεις	Patient.link	Σύνδεση με γονείς/συγγενείς
Νοσοκομείο/Μονάδα	Patient.managingOrganization	Παραπέμπουσα μονάδα

Η γεωγραφική καταγωγή αποτελεί κρίσιμο στοιχείο για τη μελέτη founder mutations και τη γεωγραφική κατανομή παραλλαγών στον ελληνικό πληθυσμό. Το σύστημα υποστηρίζει καταγραφή καταγωγής και για τους δύο γονείς, επιτρέποντας την ανάλυση πληθυσμιακών απομονώσεων (population isolates) όπως η Κρήτη και ο Πόντος.

6.1.2 Μοντελοποίηση Γενετικών Ευρημάτων

Τα γενετικά ευρήματα μοντελοποιούνται ως FHIR Observation resources ακολουθώντας το Genomics Reporting Implementation Guide. Χρησιμοποιούνται δύο τύποι Observations: ένας για το γονίδιο που εξετάστηκε (Gene Observation) και ένας για κάθε παραλλαγή που εντοπίστηκε (Variant Observation), Πίνακες 19 και 20.

Πίνακας 19 Gene Observation - Βασικά στοιχεία

Component	LOINC Code	Περιεχόμενο
Gene studied	48018-6	HGNC σύμβολο γονιδίου (π.χ. BRCA1)
Gene ID	81252-9	HGNC ID
Overall interpretation	51968-6	Positive / Negative / Inconclusive

Πίνακας 20 Variant Observation - Βασικά στοιχεία

Component	LOINC Code	Περιεχόμενο
Variant code	69548-6	Κωδικός παραλλαγής
DNA change (gHGVS)	81290-9	Γενομική θέση (π.χ. g.41193676_41193677del)
DNA change (cHGVS)	48004-6	cDNA notation (π.χ. c.5266dupC)
Protein change (pHGVS)	48005-3	Πρωτεϊνική αλλαγή (π.χ. p.Gln1756fs)
Transcript reference	51958-7	RefSeq transcript (π.χ. NM_007294.4)
Clinical significance	53037-8	ACMG classification
Allelic state (zygosity)	53034-5	Heterozygous / Homozygous / Hemizygous
Variant type	48006-1	SNV / indel / CNV
dbSNP ID	81255-2	rs identifier
Population frequency	81258-6	Allele frequency (Greek population)

6.1.3 Κλινική Σημασία και Ταξινόμηση ACMG

Η κλινική σημασία κάθε παραλλαγής ταξινομείται σύμφωνα με τις κατευθυντήριες οδηγίες του American College of Medical Genetics and Genomics (ACMG). Το σύστημα υποστηρίζει την πενταβάθμια κλίμακα ταξινόμησης με αντιστοίχιση σε κωδικούς LOINC Answer List:

Πίνακας 21 ACMG και Κλινική Σημασία

ACMG Classification	LOINC Code	Κλινική Ερμηνεία
Pathogenic (P)	LA6668-3	Παθογόνος - Αιτιολογικά συνδεδεμένη με νόσο
Likely Pathogenic (LP)	LA26332-9	Πιθανώς παθογόνος - >90% πιθανότητα
Uncertain Significance (VUS)	LA26333-7	Αβέβαιης σημασίας - Απαιτείται περαιτέρω μελέτη
Likely Benign (LB)	LA26334-5	Πιθανώς καλοήθης - >90% πιθανότητα
Benign (B)	LA6675-8	Καλοήθης - Χωρίς κλινική σημασία

Η λογική προσδιορισμού του συνολικού αποτελέσματος γενετικής ανάλυσης (Overall Interpretation) ακολουθεί σαφείς κανόνες βάσει της ταξινόμησης των ευρεθισών παραλλαγών:

Πίνακας 22 Κανόνες και Ερμηνεία

Συνθήκη	Αποτέλεσμα	SNOMED CT Code
Έστω και μία P ή LP παραλλαγή	Positive (Θετικό)	10828004
Όλες οι παραλλαγές VUS	Inconclusive (Αδιευκρίνιστο)	82334004
Μόνο B ή LB παραλλαγές ή καμία παραλλαγή	Negative (Αρνητικό)	260385009

6.1.4 Ειδικές Επεκτάσεις για τον Ελληνικό Πληθυσμό

Για την κάλυψη των ιδιαίτερων αναγκών της Ιατρικής Ακριβείας στην Ελλάδα, αναπτύχθηκαν ειδικές FHIR extensions που επιτρέπουν την καταγραφή πληροφοριών μοναδικών για τον ελληνικό πληθυσμό:

Greek Founder Mutation Extension:

Η επέκταση extension[is-greek-founder] (valueBoolean) σηματοδοτεί παραλλαγές που έχουν αναγνωριστεί ως founder mutations στον ελληνικό πληθυσμό. Οι founder mutations είναι παραλλαγές που εμφανίστηκαν σε έναν κοινό πρόγονο και διαδόθηκαν λόγω γεωγραφικής απομόνωσης και ενδογαμίας. Παραδείγματα παρουσιάζονται στον Πίνακα 23:

Πίνακας 23 Γονίδια και Παραλλαγές

Γονίδιο	Παραλλαγή	Νόσημα	Περιοχή Προέλευσης
BRCA1	c.5266dupC (p.Gln1756fs)	Κληρονομικός καρκίνος μαστού/ωοθηκών	Πανελλαδικά
BRCA1	c.5212G>A (p.Gly1738Arg)	Κληρονομικός καρκίνος μαστού/ωοθηκών	Πανελλαδικά
BRCA2	c.7806-2A>G	Κληρονομικός καρκίνος μαστού/ωοθηκών	Βόρεια Ελλάδα
BRCA1	c.5492del	Κληρονομικός καρκίνος μαστού/ωοθηκών	Κρήτη
GJB2	c.35delG	Κληρονομική απώλεια ακοής	Πανελλαδικά
HBB	c.93-21G>A (IVS-I-110)	B-Θαλασσαιμία	Πανελλαδικά

Geographic Origin Extension:

Η επέκταση extension[founder-geographic-origin] καταγράφει τη γεωγραφική περιοχή προέλευσης της παραλλαγής, επιτρέποντας τη μελέτη της γεωγραφικής κατανομής και την αναγνώριση πληθυσμιακών απομονώσεων. Οι υποστηριζόμενες περιοχές παρουσιάζονται στον Πίνακα 24:

Πίνακας 24 Γεωγραφικές Περιοχές Παραλλαγών

Κατηγορία	Περιοχές
Μεγάλες πόλεις	Αθήνα, Θεσσαλονίκη, Πάτρα, Ηράκλειο, Λάρισα, Βόλος, Ιωάννινα, Αλεξανδρούπολη
Βόρεια Ελλάδα	Φλώρινα, Κοζάνη, Καβάλα, Δράμα, Σέρρες, Κιλκίς, Ξάνθη, Κομοτηνή, Πόντος
Κεντρική Ελλάδα	Λαμία, Αμφισσα, Λιβαδειά, Χαλκίδα, Τρίκαλα, Τύρναβος
Πελοπόννησος	Τρίπολη, Καλαμάτα, Σπάρτη, Ναύπλιο, Κόρινθος
Νησιά	Κρήτη, Ρόδος, Λέσβος, Χίος, Σάμος, Κέρκυρα

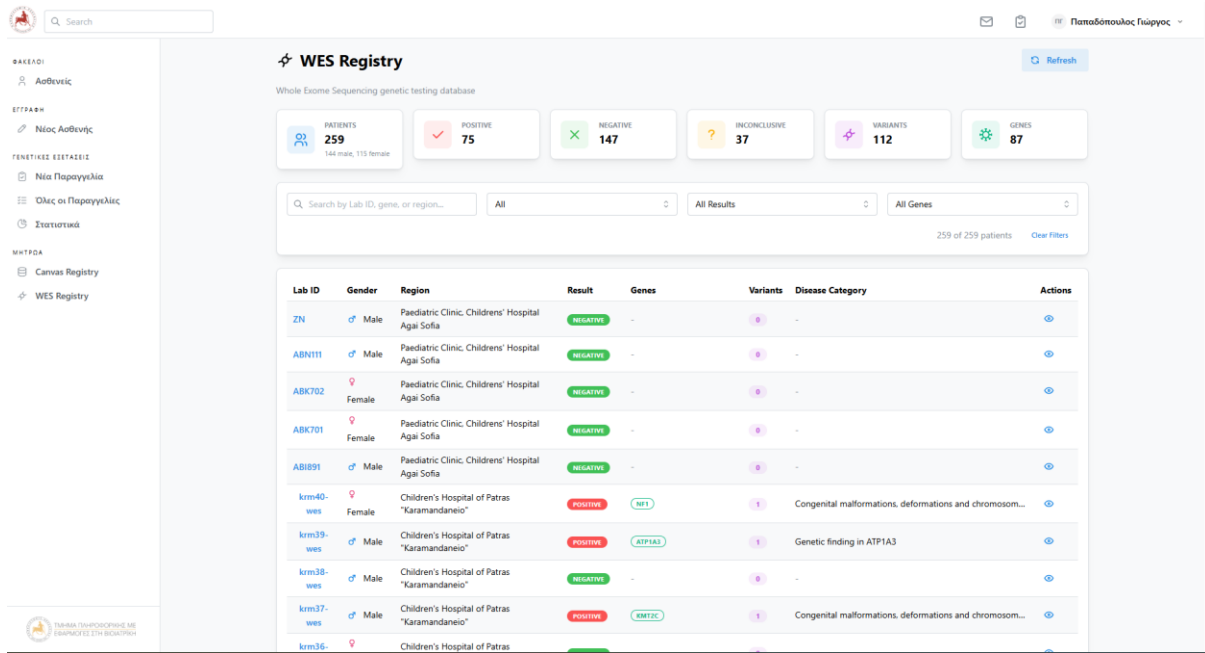
Greek Population Allele Frequency:

Για κάθε παραλλαγή καταγράφεται η συχνότητα αλληλομόρφου στον ελληνικό πληθυσμό (Greek AF), όταν αυτή είναι διαθέσιμη από τα δεδομένα του CanVaS ή άλλες πηγές. Η πληροφορία αυτή είναι κρίσιμη για την ερμηνεία παραλλαγών, καθώς παραλλαγές με υψηλή συχνότητα στον γενικό πληθυσμό είναι λιγότερο πιθανό να είναι παθογόνες.

6.2 WES Registry - Μητρώο Σπάνιων Νοσημάτων

Το WES Registry (Whole Exome Sequencing Registry) αποτελεί κλινικο-γενετικό σύστημα ολοκλήρωσης που παρακολουθεί γενετικές παραλλαγές ανακαλυπτόμενες μέσω αλληλούχισης ολόκληρου του εξωνιώματος σε νοσοκομειακό περιβάλλον. Το σύστημα επιτρέπει στους κλινικούς γενετιστές να μελετούν τη σχέση μεταξύ γενετικών παραλλαγών, νοσημάτων και κλινικών φαινοτύπων (HPO συμπτωμάτων), υποστηρίζοντας τόσο τις διαγνωστικές ροές εργασίας όσο και τις ερευνητικές δραστηριότητες, απεικόνιση διεπαφής στο Σχήμα 4.

Η κοόρτη αναπτύχθηκε στα πλαίσια της Ενότητας Εργασίας 2 (EE2) σε συνεργασία με το Πανεπιστήμιο Πατρών και περιλαμβάνει περίπου 300 ασθενείς με ποικίλα σπάνια νοσήματα, κυρίως νευρολογικά και καρδιολογικά, οι οποίοι παρέμεναν αδιάγνωστοι μέσω συμβατικών διαγνωστικών μεθόδων.



Σχήμα 4 Διεπαφή WES Registry

6.2.1 Αρχιτεκτονική FHIR Resources

Το μητρώο χρησιμοποιεί ένα συνδεδεμένο γράφημα FHIR resources για την αναπαράσταση της πλήρους κλινικο-γενετικής εικόνας κάθε ασθενούς. Η ιεραρχία των πόρων ακολουθεί τα διεθνή πρότυπα HL7 FHIR R4 για γενετικά δεδομένα, Πίνακας 25.

Πίνακας 25 Ιεραρχία FHIR Resources του WES Registry

FHIR Resource	Πηγή Δεδομένων	Σκοπός
Organization	Hospital Unit	Νοσοκομειακό τμήμα παραπομπής
Patient	Excel Individual	Δημογραφικά, σύνδεση με μονάδα
ServiceRequest	Excel Order	Εργαστηριακή παραγγελία, ΗΡΟ symptoms
Specimen	Excel Sample	Τύπος δείγματος, ημερομηνία παραλαβής
Observation (Variant)	Excel Variants	Λεπτομέρειες γενωμικής παραλλαγής, ταξινόμηση
Observation (Gene)	Excel Gene	Σύνοψη σε επίπεδο γονιδίου, σύνδεση παραλλαγών
DiagnosticReport	Generated	Σύνοψη αποτελεσμάτων WES test
Condition	Excel Disease	ICD-10/OMIM κωδικοποίηση νοσήματος

6.2.2 Πηγή και Μορφή Δεδομένων

Τα δεδομένα προέρχονται από αρχείο Excel (wes_db.xlsx) που περιέχει μία γραμμή ανά ασθενή με στήλες για δημογραφικά, κλινικά και γενετικά δεδομένα. Ο παρακάτω Πίνακας περιγράφει τη δομή των δεδομένων πηγής.

Πίνακας 26 Κατάλογος Στηλών Πηγής Δεδομένων WES

Όνομα Στήλης	Τύπος	Περιγραφή	Παράδειγμα
Unique ID	String	Μοναδικός κωδικός ασθενούς	WES-001
Sex	String	Φύλο (Male/Female)	Female
Unit	String	Νοσοκομειακό τμήμα/μονάδα	Neurology
HPO Symptoms	String	Συμπτώματα διαχωρισμένα με κόμμα	seizures, hypotonia
Result	String	Αποτέλεσμα εξέτασης	POS, NEG, INC
Gene	String	Σύμβολο επηρεαζόμενου γονιδίου	BRCA1
Classification	String	ACMG ταξινόμηση	P, LP, VUS, LP+VUS
Type of variant(s)	String	Περιγραφή τύπου παραλλαγής	missense, frameshift
Zygoty	String	Κατάσταση ζυγωτίας	Het, Hom, Hemi
Inheritance	String	Πρότυπο κληρονομικότητας	A Dom, A Rec, X Rec
Disease Category	String	ICD-10 κατηγορία νοσήματος	Diseases of nervous system
ICD-10	String	Συγκεκριμένος ICD-10 κωδικός	G11.4
OMIM code	String	Κωδικός OMIM νοσήματος	601215
Sample Type	String	Τύπος βιολογικού δείγματος	DBS, EDTA, DNA

6.2.3 Πρότυποι Κωδικοί LOINC

Το σύστημα χρησιμοποιεί τυποποιημένους κωδικούς LOINC, Πίνακας 27, για την αναγνώριση του τύπου κάθε γενετικής παρατήρησης, διασφαλίζοντας τη διαλειτουργικότητα με διεθνή συστήματα. Πίνακας 27 Κωδικοί LOINC για Γενετικές Παρατηρήσεις

LOINC Code	Display Name	Χρήση
82939-0	Whole Exome Sequencing	ServiceRequest.code
48018-6	Gene studied	Gene Observation.code
69548-6	Genetic variant assessment	Variant Observation.code
81247-9	Master HL7 genetic variant reporting panel	DiagnosticReport.code
48004-6	HGVS c. nomenclature	Variant nomenclature component
53037-8	Clinical significance	ACMG classification component
53034-5	Zygoty	Zygoty component
81258-6	Allele frequency	Population frequency component

6.2.4 Ταξινόμηση Κλινικής Σημαντικότητας (ACMG)

Οι παραλλαγές ταξινομούνται χρησιμοποιώντας το 5-βάθμιο σύστημα ACMG (American College of Medical Genetics). Τα δεδομένα πηγής χρησιμοποιούν συντομευμένους κωδικούς που αντιστοιχίζονται στις ταξινομήσεις ACMG.

Πίνακας 28 Αντιστοίχιση Ταξινομήσεων ACMG

Source Code	ACMG Classification	FHIR Code	Περιγραφή
P	Pathogenic	LA6668-3	Αιτιολογική της νόσου
LP	Likely Pathogenic	LA26332-9	Πιθανώς αιτιολογική
VUS	Uncertain Significance	LA26333-7	Άγνωστη κλινική επίπτωση
LB	Likely Benign	LA26334-5	Πιθανώς αβλαβής
B	Benign	LA6675-8	Αβλαβής παραλλαγή

Λογική Καθορισμού Αποτελέσματος

Το συνολικό αποτέλεσμα της εξέτασης καθορίζεται από τις ταξινομήσεις των παραλλαγών με χρήση κωδικών SNOMED CT, Πίνακας 29:

Πίνακας 29 Συνθήκη και Αποτελέσματα

Result	Συνθήκη	SNOMED Code
Positive	Οποιαδήποτε παραλλαγή με P ή LP	10828004 (Positive)
Inconclusive	Όλες οι παραλλαγές VUS	82334004 (Inconclusive)
Negative	Μόνο B/LB παραλλαγές ή καμία παραλλαγή	260385009 (Negative)

6.2.5 Ενσωμάτωση HPO (Human Phenotype Ontology)

Τα κλινικά συμπτώματα καταγράφονται με χρήση HPO όρων για την τυποποίηση των περιγραφών φαινοτύπου μεταξύ ασθενών. Αυτό επιτρέπει την ανάλυση συσχετίσεων φαινοτύπου-γονοτύπου, τον εντοπισμό κοορτών ασθενών με παρόμοια συμπτώματα, και την ενσωμάτωση με εργαλεία υποστήριξης διάγνωσης.

Διαδικασία Επίλυσης HPO Όρων

Κείμενο συμπτωμάτων από το Excel επιλύεται σε HPO κωδικούς μέσω του JAX Ontology API, Πίνακας 30. Η διαδικασία περιλαμβάνει: (1) Ανάλυση κειμένου συμπτωμάτων σε όρους, (2) Αναζήτηση κάθε όρου στο JAX API, (3) Αποθήκευση σε cache για αποφυγή επαναλαμβανόμενων κλήσεων, (4) Αποθήκευση στο ServiceRequest.reasonCode.

Πίνακας 30 Περιγραφή Συμπτώματος

Περιγραφή Συμπτώματος	HPO Code	HPO Term
Σπασμοί, επιληπτικές κρίσεις	HP:0001250	Seizure
Νοητική υστέρηση	HP:0001249	Intellectual disability
Μυϊκή υποτονία	HP:0001252	Muscular hypotonia
Αταξία	HP:0001251	Ataxia

Καρδιομυοπάθεια	HP:0001638	Cardiomyopathy
Μικροκεφαλία	HP:0000252	Microcephaly
Σπαστικότητα	HP:0001257	Spasticity
Αναπτυξιακή καθυστέρηση	HP:0001263	Global developmental delay

6.2.6 Ταξινόμηση Τύπων Παραλλαγών

Οι τύποι παραλλαγών πηγής, Πίνακας 31, αντιστοιχίζονται σε τυποποιημένες κατηγορίες FHIR για ομοίμορφη αναπαράσταση.

Πίνακας 31 Αντιστοίχιση Τύπων Παραλλαγών

Source Type	FHIR Variant Type	Category	Περιγραφή
missense	SNV	missense	Αντικατάσταση αμινοξέος
nonsense	SNV	nonsense	Πρόωρο κωδικόνιο τερματισμού
frameshift (dup)	indel	Frameshift (in)	Μετατόπιση πλαισίου από διπλασιασμό
frameshift (del)	indel	frameshift (del)	Μετατόπιση πλαισίου από διαγραφή
CNV / large deletion	CNV	other	Παραλλαγή αριθμού αντιγράφων
splice/intron	SNV	non-coding intron	Παραλλαγή θέσης ματίσματος
inframe (del/in)	indel	inframe (del/in)	In-frame εισαγωγή/διαγραφή

6.2.7 Κατηγοριοποίηση Νοσημάτων κατά ICD-10

Τα νοσήματα κατηγοριοποιούνται σύμφωνα με τα κεφάλαια του ICD-10, επιτρέποντας τη στατιστική ανάλυση ανά σύστημα οργάνων. Οι κύριες κατηγορίες που εμφανίζονται στο μητρώο περιλαμβάνουν:

Πίνακας 32 Κατηγορίες ICD-10 στο WES Registry

ICD-10 Range	Κατηγορία Νοσημάτων	% στο Μητρώο
G00-G99	Νοσήματα νευρικού συστήματος	~45%
I00-I99	Νοσήματα κυκλοφορικού συστήματος	~20%
Q00-Q99	Συγγενείς ανωμαλίες, παραμορφώσεις και χρωμοσωμικές ανωμαλίες	~15%
E00-E89	Ενδοκρινικά, διατροφικά και μεταβολικά νοσήματα	~10%
M00-M99	Νοσήματα μυοσκελετικού συστήματος και συνδετικού ιστού	~5%

F01-F99	Ψυχικές, συμπεριφορικές και νευροαναπτυξιακές διαταραχές	~3%
Άλλες	Διάφορες κατηγορίες (H, D, L, N κ.λπ.)	~2%

6.2.8 Χειρισμός Σύνθετης Ετεροζυγωτίας

Για υπολειπόμενα (autosomal recessive) νοσήματα, το μητρώο υποστηρίζει τον χειρισμό περιπτώσεων σύνθετης ετεροζυγωτίας (compound heterozygosity), όπου δύο διαφορετικές παραλλαγές στο ίδιο γονίδιο συμβάλλουν στη νόσο.

Ανίχνευση Σύνθετης Ετεροζυγωτίας

Η σύνθετη ετεροζυγωτία υποδεικνύεται όταν: το πεδίο Quantity περιέχει "+" (π.χ. "1+1") ή το πεδίο Classification περιέχει "+" (π.χ. "LP+VUS"). Σε αυτές τις περιπτώσεις, το σύστημα αναλύει τις ταξινομήσεις και τους τύπους παραλλαγών και δημιουργεί ξεχωριστά Variant Observations για κάθε παραλλαγή με κατάλληλη σύνδεση μεταξύ τους.

6.2.9 Ζυγωτία και Πρότυπα Κληρονομικότητας

Οι Πίνακες 33 και 34 παρουσιάζουν την αντιστοίχιση Ζυγωτίας καθώς και την αντιστοίχιση των προτύπων κληρονομικότητας.

Πίνακας 33 Αντιστοίχιση Ζυγωτίας

Source Value	FHIR Display	Περιγραφή
Het	Heterozygous	Ένα αλληλόμορφο επηρεασμένο
Hom	Homozygous	Και τα δύο αλληλόμορφα επηρεασμένα
Hemi	Hemizygous	Ένα αλληλόμορφο (X-linked σε άνδρες)

Πίνακας 34 Αντιστοίχιση Προτύπων Κληρονομικότητας

Source Value	FHIR Value	Περιγραφή
A Rec	Autosomal Recessive	Αυτοσωματική υπολειπόμενη
A Dom	Autosomal Dominant	Αυτοσωματική επικρατούσα
X Rec / X-linked Rec	X-linked Recessive	Φυλοσύνδετη υπολειπόμενη
X Dom / X-linked Dom	X-linked Dominant	Φυλοσύνδετη επικρατούσα
Mit	Mitochondrial	Μιτοχονδριακή κληρονομικότητα

6.2.10 Αγωγός Δεδομένων (Data Pipeline)

Ο αγωγός δεδομένων του WES Registry μετασχηματίζει γενετικά δεδομένα από τη βάση Excel σε FHIR R4 resources. Η διαδικασία περιλαμβάνει τα ακόλουθα στάδια:

1. Φόρτωση ICD-10 Code System: Τα display names για κωδικούς ICD-10 φορτώνονται από το πακέτο definitions για παροχή αναγνώσιμων ονομασιών νοσημάτων.
2. Ανάλυση αρχείου Excel: Ανάγνωση του αρχείου wes_db.xlsx με τη βιβλιοθήκη xlsx.

3. Επίλυση HPO όρων: Για κάθε συμπτώματα αναζήτηση στο JAX Ontology API με rate limiting (100ms καθυστέρηση) και caching αποτελεσμάτων.
4. Δημιουργία FHIR Bundles: Για κάθε γραμμή δημιουργείται ξεχωριστό transaction bundle με όλους τους σχετικούς πόρους.
5. Εισαγωγή στο Medplum: Τα bundles εισάγονται μέσω POST στο FHIR endpoint με idempotent μηχανισμό για Organizations.

Δομή Bundle ανά Ασθενή

Κάθε ασθενής αντιστοιχεί σε ένα FHIR transaction bundle με την ακόλουθη δομή (οι πόροι ταξινομούνται για ικανοποίηση εξαρτήσεων αναφορών, Πίνακας 35):

Πίνακας 35 Πόροι και Εξαρτήσεις

Σειρά	Resource	Εξαρτήσεις
1	Organization	Καμία - conditional create
2	Patient	References Organization
3	Specimen	References Patient, ServiceRequest
4	ServiceRequest	References Patient, Specimen, Organization
5	Observation (Variant)	References Patient (0-N ανά παραλλαγή)
6	Observation (Gene)	References Patient, hasMember → Variants
7	Condition	References Patient, evidence → Gene, Variants
8	DiagnosticReport	References Patient, ServiceRequest, result → Gene Observations

6.2.11 Σύστημα Tags για Απομόνωση Πόρων

Οι πόροι WES χρησιμοποιούν FHIR meta.tag για αποτελεσματικό querying χωρίς ανάμιξη με άλλους ασθενείς της πλατφόρμας.

Πίνακας 36 Σύστημα Tags WES Registry

Resource Type	Tag Code	Tag Display
Patient	wes-registry	WES Registry Patient
Observation (Variant)	wes-variant	WES Variant Observation
Observation (Gene)	wes-gene	WES Gene Observation
DiagnosticReport	wes-report	WES Diagnostic Report

Tag System URI: <http://wes-database.example.org/tags>

6.2.12 Στατιστικά και Δείκτες Μητρώου

Το μητρώο παρέχει αυτόματη παραγωγή στατιστικών και δεικτών απόδοσης που είναι κρίσιμοι για την αξιολόγηση της διαγνωστικής διαδικασίας.

Δείκτης	Περιγραφή	Τρόπος Υπολογισμού
Diagnostic Yield	Ποσοστό θετικών αποτελεσμάτων	Positive / Total × 100%
VUS Rate	Ποσοστό αδιευκρίνιστων παραλλαγών	Inconclusive / Total × 100%
Negative Rate	Ποσοστό αρνητικών αποτελεσμάτων	Negative / Total × 100%
Gene Distribution	Κατανομή ευρημάτων ανά γονίδιο	Count per gene
Phenotype-Genotype	Συσχετίσεις HPO-Gene	Co-occurrence analysis
Time to Diagnosis	Χρόνος μέχρι διάγνωση	Diagnosis date - Onset date

6.3 CanVaS Registry - Μητρώο Κληρονομικού Καρκίνου

Το CanVaS (Cancer Variation reSource) αποτελεί εθνικό μητρώο γενετικών παραλλαγών από Έλληνες με προδιάθεση σε καρκίνο. Αναπτύχθηκε από το Εργαστήριο Μοριακής Διαγνωστικής του ΕΚΕΦΕ «Δημόκριτος» και δημοσιεύθηκε στο περιοδικό Human Mutation¹⁴ (Kalfakakou et al., 2021). Στα πλαίσια του έργου TAEDR-0539180, τα δεδομένα του CanVaS ενσωματώθηκαν πλήρως στην πλατφόρμα FHIR R4 με διατήρηση όλης της κλινικής και γενετικής πληροφορίας, καθώς και προσθήκη δυνατοτήτων διαλειτουργικότητας σύμφωνα με το EHDS.

Η σημασία των εθνικών μητρώων γενετικών παραλλαγών είναι ιδιαίτερα κρίσιμη για πληθυσμούς που υποεκπροσωπούνται σε μεγάλες διεθνείς βάσεις δεδομένων όπως το gnomAD. Ο ελληνικός πληθυσμός παρουσιάζει μοναδική γενετική σύνθεση, με έντονη γενετική ετερογένεια αλλά και ισχυρά φαινόμενα founder effect λόγω ιστορικής γεωγραφικής απομόνωσης ορεινών και νησιωτικών κοινοτήτων. Το CanVaS καταγράφει αυτή τη μοναδική γενετική ποικιλομορφία, επιτρέποντας τον υπολογισμό πληθυσμιακά-ειδικών συχνοτήτων αλληλομόρφων και την αναγνώριση σπάνιων παραλλαγών που σχετίζονται με κίνδυνο νόσου, απεικόνιση της διεπαφής στο Σχήμα 5.

The screenshot displays the Canvas Registry interface. At the top, there is a search bar and a 'Refresh' button. Below this, a summary shows 7,372 patients and 23,719 variants. A search bar is provided for finding patients by Canvas ID or Lab ID. The main table lists patient records with the following columns: Canvas ID, Gender, Region, Variants, Genes, Diseases, Founder, and Actions. The table contains 15 rows of data, including patient IDs, genders, regions (e.g., Lesbos, Mikra Asia & Thrace, Pyrgos, Piraeus, 'Elliniko, Arcadia', Corfu, 'Antalya, Turkey', Lefkada, 'Fostaina, Achaia'), variant counts, associated genes (e.g., CHEK2, MSH2, CHEK2, ERCC1, ERCC2, ERCC3, ERCC4, ERCC5, ERCC6, ERCC7, ERCC8, ERCC9, ERCC10, ERCC11, ERCC12, ERCC13, ERCC14, ERCC15, ERCC16, ERCC17, ERCC18, ERCC19, ERCC20, ERCC21, ERCC22, ERCC23, ERCC24, ERCC25, ERCC26, ERCC27, ERCC28, ERCC29, ERCC30, ERCC31, ERCC32, ERCC33, ERCC34, ERCC35, ERCC36, ERCC37, ERCC38, ERCC39, ERCC40, ERCC41, ERCC42, ERCC43, ERCC44, ERCC45, ERCC46, ERCC47, ERCC48, ERCC49, ERCC50, ERCC51, ERCC52, ERCC53, ERCC54, ERCC55, ERCC56, ERCC57, ERCC58, ERCC59, ERCC60, ERCC61, ERCC62, ERCC63, ERCC64, ERCC65, ERCC66, ERCC67, ERCC68, ERCC69, ERCC70, ERCC71, ERCC72, ERCC73, ERCC74, ERCC75, ERCC76, ERCC77, ERCC78, ERCC79, ERCC80, ERCC81, ERCC82, ERCC83, ERCC84, ERCC85, ERCC86, ERCC87, ERCC88, ERCC89, ERCC90, ERCC91, ERCC92, ERCC93, ERCC94, ERCC95, ERCC96, ERCC97, ERCC98, ERCC99, ERCC100), and disease counts.

Σχήμα 5 Διεπαφή Canvas Registry

¹⁴ <https://doi.org/10.1002/humu.24249> Digital Object Identifier (DOI)

6.3.1 Προέλευση και Δομή Δεδομένων

Τα δεδομένα του CanVaS προέρχονται από γενετικό έλεγχο germline DNA ασθενών που παραπέμφθηκαν λόγω ατομικού ή οικογενειακού ιστορικού κακοήθειας στο Εργαστήριο Μοριακής Διαγνωστικής του ΕΚΕΦΕ «Δημόκριτος» κατά την περίοδο 1999-2019. Όλοι οι συμμετέχοντες έδωσαν γραπτή ενημερωμένη συγκατάθεση, επιτρέποντας τη χρήση ανωνυμοποιημένων δεδομένων για ερευνητικούς σκοπούς. Οι ασθενείς ερωτήθηκαν για το ατομικό και οικογενειακό τους ιστορικό καρκίνου, καθώς και για τη γεωγραφική καταγωγή των γονέων τους.

Μεθοδολογία Γενετικού Ελέγχου

Ο γενετικός έλεγχος εξελίχθηκε διαχρονικά ακολουθώντας τις τεχνολογικές εξελίξεις: κατά την περίοδο 1999-2012 χρησιμοποιήθηκε αποκλειστικά αλληλούχιση κατά Sanger, ενώ μετά το 2012 εισήχθησαν τα NGS multigene panels, εκτός εάν ο ασθενής είχε σαφή κλινική διάγνωση γενετικού συνδρόμου οπότε εφαρμόζοταν στοχευμένη αλληλούχιση του υποψήφιου γονιδίου. Επιπλέον, σε όλες τις αναλύσεις εφαρμόστηκε συμπληρωματικά η τεχνική MLPA (Multiplex Ligation-dependent Probe Amplification) για την ανίχνευση μεγάλων γενωμικών αναδιατάξεων (deletions/duplications) που καλύπτουν ένα ή περισσότερα εξώνια.

Βασικά στατιστικά CanVaS Registry:

Πίνακας 37 Μετρικές και Τιμές του CanVaS

Μετρική	Τιμή
Συνολικά άτομα	7.363
Γυναίκες	6.271 (85,17%)
Άνδρες	1.092 (14,83%)
Γονίδια προδιάθεσης	97
Συνολικές καταχωρήσεις παραλλαγών	22.089
Μοναδικές παραλλαγές	7.968
Pathogenic/Likely Pathogenic (P/LP)	2.204 (9,98%)
Variants of Uncertain Significance (VUS)	10.696 (48,42%)
Benign/Likely Benign (B/LB)	9.189 (41,60%)

6.3.2 Γονίδια Προδιάθεσης στον Καρκίνο

Το CanVaS περιλαμβάνει γενετικά δεδομένα από 97 γονίδια που σχετίζονται με κληρονομική προδιάθεση στον καρκίνο. Τα γονίδια αυτά κατηγοριοποιούνται με βάση τη διεισδυτικότητα (penetrance) και την κλινική τους αξιοποιησιμότητα σύμφωνα με τις κατευθυντήριες οδηγίες ACMG και NCCN.

Κατηγοριοποίηση Γονιδίων κατά Διεισδυτικότητα:

Πίνακας 38 Παραδείγματα Κατηγοριοποίησης Γονιδίων

Κατηγορία Διεισδυτικότητας	Αριθμός Γονιδίων	Παραδείγματα	% P/LP Παραλλαγών
Υψηλής διεισδυτικότητας	44	BRCA1, BRCA2, APC, MLH1, MSH2, TP53, VHL, RET	82,5%
Μέτριας διεισδυτικότητας	10	ATM, CHEK2, PALB2, BRIP1, RAD51C, RAD51D	12,0%

Χαμηλής/αβέβαιης διεισδυτικότητας	43	BAP1, BARD1, NBN, WRN, EGFR, KIT	5,5%
-----------------------------------	----	----------------------------------	------

Αξιοσημείωτο είναι ότι 1.643 (74,6%) παθογόνες/πιθανώς παθογόνες παραλλαγές αφορούν κλινικά αξιοποιήσιμα γονίδια που περιλαμβάνονται στη λίστα ACMG SF v2.0 των 59 γονιδίων για τα οποία συνιστάται η αναφορά τυχαίων ευρημάτων. Αυτό υπογραμμίζει τη σημασία του μητρώου για τη λήψη κλινικών αποφάσεων, καθώς η αναγνώριση παθογόνων παραλλαγών σε αυτά τα γονίδια μπορεί να οδηγήσει σε συγκεκριμένες παρεμβάσεις πρόληψης ή θεραπείας.

Κύρια Κλινικά Σύνδρομα και Σχετιζόμενα Γονίδια:

Πίνακας 39 Συσχέτιση σύνδρομου και γονιδίου

Σύνδρομο/Νόσος	Γονίδια	Κληρονομικότητα
Hereditary Breast and Ovarian Cancer (HBOC)	BRCA1, BRCA2, PALB2	AD
Lynch Syndrome	MLH1, MSH2, MSH6, PMS2, EPCAM	AD
Familial Adenomatous Polyposis (FAP)	APC	AD
Li-Fraumeni Syndrome	TP53	AD
Von Hippel-Lindau Syndrome	VHL	AD
Multiple Endocrine Neoplasia Type 2	RET	AD
Peutz-Jeghers Syndrome	STK11	AD
Cowden Syndrome	PTEN	AD

AD = Αυτοσωμική Επικρατής, AR = Αυτοσωμική Υπολειπόμενη

6.3.3 Ελληνικές Founder Mutations

Ιδιαίτερη σημασία για την Ιατρική Ακριβείας στην Ελλάδα έχει η τεκμηρίωση των founder mutations στον ελληνικό πληθυσμό. Οι founder mutations είναι παθογόνες παραλλαγές που εμφανίστηκαν σε έναν κοινό πρόγονο και διαδόθηκαν σε επόμενες γενεές λόγω γεωγραφικής απομόνωσης και ενδογαμίας. Η αναγνώριση αυτών των παραλλαγών έχει άμεσες κλινικές εφαρμογές, καθώς επιτρέπει τη στοχευμένη γενετική εξέταση και τη βελτιστοποίηση των διαγνωστικών πρωτοκόλλων για τον ελληνικό πληθυσμό.

Το CanVaS καταγράφει τη γεωγραφική κατανομή κάθε παραλλαγής σε επίπεδο χωριού/πόλης καταγωγής και των δύο γονέων, επιτρέποντας την αναγνώριση πληθυσμιακών απομονώσεων και τη μελέτη των founder effects. Αυτή η πληροφορία είναι κρίσιμη για την κατανόηση της γενετικής δομής του ελληνικού πληθυσμού και για την ερμηνεία σπάνιων παραλλαγών.

Πίνακας 40 Τεκμηριωμένες Ελληνικές Founder Mutations

Γονίδιο	Παραλλαγή (HGVS)	Σχετιζόμενη Νόσος	Περιοχή Προέλευσης
BRCA1	NM_007294.3:c.5266dupC (p.Gln1756Profs)	HBOC	Πανελλαδικά
BRCA1	NM_007294.3:c.5212G>A (p.Gly1738Arg)	HBOC (53 οικογένειες)	Πανελλαδικά

BRCA1	Exon 24 deletion	HBOC	Πολλαπλές περιοχές
BRCA1	NM_007294.3:c.5492del	HBOC	Κρήτη
BRCA2	NM_000059.3:c.7806-2A>T	HBOC	Κρήτη
BRCA2	NM_000059.3:c.6842-2675_7008-5558del	HBOC	Κρήτη
GJB2	NM_004004.6:c.35delG	Κώφωση	Πανελλαδικά

Μελέτη περίπτωσης - Κρητικές απομονώσεις:

Χαρακτηριστικό παράδειγμα του founder effect αποτελεί η μελέτη founder mutations στις κρητικές γενετικές απομονώσεις (Apostolou et al., 2020). Τρεις παθογόνες παραλλαγές μοναδικές για τον κρητικό υποπληθυσμό - η BRCA1 c.5492del και οι BRCA2 c.7806-2A>T και c.6842-2675_7008-5558del - ευθύνονται για το 48% των P/LP παραλλαγών σε BRCA1/BRCA2 μεταξύ ασθενών κρητικής καταγωγής με καρκίνο μαστού ή ωοθηκών. Αυτή η συγκέντρωση founder mutations σε συγκεκριμένη γεωγραφική περιοχή έχει σημαντικές προεκτάσεις για τον σχεδιασμό στοχευμένων προγραμμάτων πρόληψης.

Παράδειγμα ταξινόμησης founder mutation:

Η παραλλαγή BRCA1 c.5212G>A (p.Gly1738Arg) αποτελεί χαρακτηριστικό παράδειγμα του πώς η μελέτη πληθυσμιακά-ειδικών δεδομένων συμβάλλει στην ταξινόμηση παραλλαγών. Ως missense παραλλαγή, η παθογονικότητά της ήταν δύσκολο να τεκμηριωθεί με κλασικά κριτήρια. Ωστόσο, η επανειλημμένη εμφάνισή της σε 53 ελληνικές οικογένειες με καρκίνο μαστού/ωοθηκών, σε συνδυασμό με αναλύσεις συνδιαχωρισμού (co-segregation), επέτρεψε τη σωστή ταξινόμησή της ως παθογόνος παραλλαγή. Αυτή η πληροφορία είναι κρίσιμη όχι μόνο για ασθενείς στην Ελλάδα αλλά και για την ελληνική διασπορά παγκοσμίως.

6.3.4 Ενσωμάτωση Φαινοτυπικών Δεδομένων

Κρίσιμο χαρακτηριστικό του CanVaS που το διαφοροποιεί από άλλες βάσεις γενετικών παραλλαγών είναι η λεπτομερής καταγραφή φαινοτυπικών δεδομένων. Σε αντίθεση με βάσεις όπως το ClinVar που επικεντρώνονται κυρίως στην παραλλαγή, το CanVaS διασυνδέει κάθε γενετικό εύρημα με πλούσια κλινική πληροφορία του φορέα, επιτρέποντας αναλύσεις συσχέτισης γονοτύπου-φαινοτύπου, Πίνακας 41.

Πίνακας 41 Κατανομή Φαινοτύπων στο CanVaS

Φαινότυπος	Αριθμός Ατόμων	Ποσοστό	Μέση Ηλικία Διάγνωσης
Καρκίνος μαστού (γυναίκες)	3.867	52,52%	46,8 ± 12,1 έτη
Καρκίνος μαστού (άνδρες)	103	1,40%	62,0 ± 13,4 έτη
Καρκίνος ωοθηκών	1.017	13,81%	53,1 ± 13,0 έτη
Καρκίνος παχέος / Πολύποδες	510	6,93%	50,6 ± 13,9 έτη
Καρκίνος παγκρέατος	158	2,15%	60,1 ± 12,2 έτη
Καρκίνος θυρεοειδούς	131	1,78%	46,4 ± 15,0 έτη

Καρκίνος ενδομητρίου	102	1,39%	52,1 ± 12,8 έτη
Λοιποί καρκινικοί φαινότυποι	857	11,64%	—
Μη-καρκινικά σύνδρομα	317	4,30%	—
Υγιείς συγγενείς	1.362	18,50%	—

Για κάθε φαινότυπο καταγράφονται λεπτομερή κλινικά δεδομένα, συμπεριλαμβανομένων της μορφολογίας του όγκου, του σταδίου και βαθμού κακοήθειας, της κατάστασης υποδοχέων ορμονών (ER, PR, HER2) για καρκίνο μαστού, της κατάστασης μικροδορυφορικής αστάθειας (MSI) για καρκίνο παχέος εντέρου, του Gleason score για καρκίνο προστάτη, και του οικογενειακού ιστορικού σχετικών κακοηθειών. Αυτή η πληροφορία επιτρέπει εκτενείς αναλύσεις φαινοτυπικής ετερογένειας και συσχετίσεων γονοτύπου-φαινοτύπου.

6.3.5 Τεχνική Ενσωμάτωση στο FHIR

Η μεταφορά των δεδομένων του CanVaS από το πρωτότυπο LOVD-based σύστημα στην πλατφόρμα FHIR R4 πραγματοποιήθηκε μέσω εξειδικευμένου ETL (Extract-Transform-Load) pipeline. Το pipeline μετατρέπει κάθε εγγραφή του LOVD σε τυποποιημένους FHIR resources, διατηρώντας πλήρως τη σημασιολογική πληροφορία και προσθέτοντας δυνατότητες διαλειτουργικότητας, Πίνακας 42.

Πίνακας 42 FHIR Resource Mapping ανά Ασθενή

FHIR Resource	Περιεχόμενο από LOVD	Πλήθος	LOINC Code
Patient	Individual: δημογραφικά, γεωγραφική καταγωγή, συγγένεια γονέων, οικογενειακές συνδέσεις	1	—
Observation (Variant)	Variants_On_Genome: γενομική θέση, HGVS notation, ACMG classification, Greek AF	N	69548-6
Observation (Gene)	Genes: γονιδιακή σύνοψη, σύνδεση με παραλλαγές	M	48018-6
Condition	Diseases: διάγνωση συνδεδεμένη με γενετική τεκμηρίωση, OMIM coding	K	—
ServiceRequest	Screenings: εργαστηριακή διαδικασία, τεχνική αλληλούχισης	1	—
DiagnosticReport	Registry summary: συγκεντρωτικά αποτελέσματα ανά ασθενή	1	81247-9

Custom FHIR Extensions για Ελληνικά Πληθυσμιακά Δεδομένα

Για την πλήρη αναπαράσταση της πληθυσμιακά-ειδικής πληροφορίας του CanVaS, αναπτύχθηκαν custom FHIR extensions με base URL:

<http://canvas.rrp.demokritos.gr/fhir/StructureDefinition>

Πίνακας 43 URL Extensions

Extension URL (σύντομη ονομασία)	Τύπος Τιμής	Περιγραφή
is-greek-founder	boolean	Σήμανση ελληνικής founder mutation
greek-frequency	decimal	Συχνότητα αλληλομόρφου στον ελληνικό πληθυσμό

greek-geographic-origin	string	Γεωγραφική καταγωγή ασθενούς (γονέας 1)
greek-geographic-origin-2	string	Γεωγραφική καταγωγή ασθενούς (γονέας 2)
population-origin	string	Εθνική/πληθυσμιακή ομάδα
consanguinity	boolean	Κατάσταση συγγένειας γονέων
founder-geographic-origin	string	Περιοχή προέλευσης founder variant
canvas-individual-id	integer	Αρχικό LOVD ID για ιχνηλασιμότητα

Συστήματα Ταυτοποίησης (Identifier Systems)

Για τη μοναδική ταυτοποίηση και την αποφυγή διπλότυπων εγγραφών, το σύστημα χρησιμοποιεί 8 identifier systems που επιτρέπουν την αναζήτηση και τη διασύνδεση resources:

Πίνακας 44 Identifier System

Identifier System URI	Παράδειγμα Τιμής	Χρήση σε Resource
http://canvas.rrp.demokritos.gr/individuals	12345	Patient
http://canvas.rrp.demokritos.gr/lab-ids	100fam	Patient
http://canvas.rrp.demokritos.gr/variants	54321	Observation (Variant)
http://canvas.rrp.demokritos.gr/variant-dbids	BRCA2_000010	Observation (Variant)
http://canvas.rrp.demokritos.gr/diseases	100	Condition
http://www.omim.org	113705	Condition
http://www.genenames.org	BRCA1	Observation (Gene)

Σύστημα Tags για Απομόνωση Resources

Τα CanVaS resources χρησιμοποιούν FHIR meta.tag για αποδοτική αναζήτηση χωρίς ανάμειξη με άλλους ασθενείς της πλατφόρμας. Το tag system είναι: <http://canvas.rrp.demokritos.gr/tags>

Πίνακας 45 Tag System

Resource Type	Tag Code	Tag Display
Patient	canvas-registry	Canvas Registry Patient
Observation (Variant)	canvas-variant	Canvas Registry Variant
Observation (Gene)	canvas-gene	Canvas Registry Gene
DiagnosticReport	canvas-report	Canvas Registry Report

6.3.6 Κλινική Αξιοποίηση και Μελλοντικές Προοπτικές

Η ενσωμάτωση του CanVaS στην πλατφόρμα E-ΗΦΥ παρέχει άμεσες κλινικές εφαρμογές για την Ιατρική Ακριβείας στην Ελλάδα. Οι κλινικοί ιατροί μπορούν να αξιοποιούν τις πληθυσμιακά-ειδικές συχνότητες για την ερμηνεία παραλλαγών, ενώ η πληροφορία για τις ελληνικές founder mutations επιτρέπει τη βελτιστοποίηση των διαγνωστικών πρωτοκόλλων.

Κύριες Κλινικές Εφαρμογές:

- Ερμηνεία VUS (Variants of Uncertain Significance) με βάση πληθυσμιακά-ειδικά δεδομένα συχνότητας και φαινοτύπου
- Στοχευμένος γενετικός έλεγχος για γνωστές ελληνικές founder mutations ανάλογα με τη γεωγραφική καταγωγή
- Cascade testing σε συγγενείς φορέων παθογόνων παραλλαγών με διασύνδεση οικογενειακών εγγραφών
- Αναλύσεις συσχέτισης γονοτύπου-φαινοτύπου για φαινοτυπική ετερογένεια
- Υποστήριξη κλινικών αποφάσεων για πρόληψη και θεραπεία βάσει γενετικού προφίλ

Μελλοντικές Κατευθύνσεις: Η ενσωμάτωση του CanVaS αποτελεί το πρώτο βήμα μιας συντονισμένης προσπάθειας για τη δημιουργία εθνικού μητρώου γενετικών παραλλαγών. Μακροπρόθεσμος στόχος είναι η συνεργασία με άλλα δημόσια και ιδιωτικά διαγνωστικά εργαστήρια για τη συγκέντρωση γενετικών και φαινοτυπικών δεδομένων που συμβάλλουν στη γενετική προδιάθεση καρκίνου μεταξύ Ελλήνων. Η διαλειτουργικότητα μέσω FHIR και η εναρμόνιση με το EHDS θα επιτρέψουν τη διασυνοριακή ανταλλαγή δεδομένων και τη συμμετοχή σε ευρωπαϊκές πρωτοβουλίες Ιατρικής Ακριβείας.

6.4 OMOP CDM και Αναλυτικό Επίπεδο

Το αναλυτικό επίπεδο του συστήματος βασίζεται στο OMOP Common Data Model (CDM) με την επέκταση Genomic CDM, παρέχοντας δυνατότητες για δευτερογενή ανάλυση δεδομένων, συμμετοχή σε federated δίκτυα έρευνας, και διεξαγωγή αποκεντρωμένων αναλύσεων χωρίς μεταφορά πρωτογενών δεδομένων.

6.4.1 Διαδικασία ETL FHIR → OMOP

Η μεταφορά δεδομένων από το επιχειρησιακό επίπεδο (FHIR) στο αναλυτικό επίπεδο (OMOP) πραγματοποιείται μέσω διαδικασιών ETL (Extract-Transform-Load) που εκτελούνται περιοδικά. Η διαδικασία περιλαμβάνει τρία στάδια:

Extract: Ανάκτηση νέων ή τροποποιημένων FHIR resources από τον FHIR Server με χρήση του `_lastUpdated` search parameter για incremental updates.

Transform: Μετασηματισμός των resources σύμφωνα με τους κανόνες mapping, αντιστοίχιση κωδικών (ICD-10, SNOMED, LOINC, HGNC) σε OMOP concept IDs, και εφαρμογή ψευδωνυμοποίησης για προστασία προσωπικών δεδομένων.

Load: Εισαγωγή στους πίνακες OMOP με σωστή διαχείριση foreign keys και διατήρηση referential integrity.

Πίνακας 46 Αντιστοίχιση FHIR Resources σε OMOP Tables

FHIR Resource	OMOP Table(s)	Genomic CDM Table
Patient	PERSON, LOCATION	-
Condition	CONDITION_OCCURRENCE	-
Observation (Gene)	MEASUREMENT	TARGET_GENE

Observation (Variant)	MEASUREMENT	VARIANT_OCCURRENCE, VARIANT_ANNOTATION
DiagnosticReport	PROCEDURE_OCCURRENCE	-
ServiceRequest	PROCEDURE_OCCURRENCE	-
Specimen	SPECIMEN	-
RiskAssessment (PRS)	MEASUREMENT	MEASUREMENT (genomic extension)

6.4.2 Genomic CDM Extension Tables

Η επέκταση Genomic CDM του OHDSI προσθέτει εξειδικευμένους πίνακες για την αποθήκευση γενετικών δεδομένων με τρόπο κατάλληλο για αναλυτικά ερωτήματα:

TARGET_GENE - Γονίδια που εξετάστηκαν:

Πίνακας 47 Γονίδια προς Εξέταση

Πεδίο	Τύπος	Περιγραφή
target_gene_id	INTEGER	Primary key
person_id	INTEGER	Foreign key → PERSON
gene_concept_id	INTEGER	OMOP concept για το γονίδιο
gene_source_value	VARCHAR	HGNC symbol (π.χ. BRCA1)
test_type_concept_id	INTEGER	Τύπος εξέτασης (WES, panel, κλπ.)
result_concept_id	INTEGER	Αποτέλεσμα (positive, negative, VUS)
test_date	DATE	Ημερομηνία εξέτασης

VARIANT_OCCURRENCE - Παραλλαγές που εντοπίστηκαν:

Πίνακας 48 Πεδία και Παραλλαγές

Πεδίο	Τύπος	Περιγραφή
variant_occurrence_id	INTEGER	Primary key
person_id	INTEGER	Foreign key → PERSON
target_gene_id	INTEGER	Foreign key → TARGET_GENE
reference_sequence	VARCHAR	RefSeq transcript (π.χ. NM_007294.4)
hgvs_c	VARCHAR	cDNA notation (π.χ. c.5266dupC)
hgvs_p	VARCHAR	Protein notation (π.χ. p.Gln1756fs)
chromosome	VARCHAR	Χρωμόσωμα (1-22, X, Y)
position	INTEGER	Γενομική θέση (GRCh38)
ref_allele	VARCHAR	Reference allele
alt_allele	VARCHAR	Alternate allele

zygosity_concept_id	INTEGER	Ζυγωτία
allele_frequency	FLOAT	Συχνότητα στον ελληνικό πληθυσμό
is_greek_founder	BOOLEAN	Founder mutation flag

VARIANT_ANNOTATION - Ερμηνείες παραλλαγών:

Πίνακας 49 Ερμηνεία Παραλλαγών

Πεδίο	Τύπος	Περιγραφή
variant_annotation_id	INTEGER	Primary key
variant_occurrence_id	INTEGER	Foreign key → VARIANT_OCCURRENCE
classification_concept_id	INTEGER	ACMG classification concept
classification_source	VARCHAR	Πηγή (ClinVar, lab, literature)
evidence_codes	VARCHAR	ACMG evidence codes (PP1, PS3, κλπ.)
annotation_date	DATE	Ημερομηνία ταξινόμησης
curator	VARCHAR	Υπεύθυνος ταξινόμησης

6.4.3 Federated Analytics Capability

Η υιοθέτηση του OMOP CDM επιτρέπει τη συμμετοχή σε federated δίκτυα έρευνας όπως το OHDSI (Observational Health Data Sciences and Informatics). Στο federated μοντέλο, κάθε ίδρυμα διατηρεί τα δεδομένα του τοπικά σε OMOP format, και οι αναλύσεις εκτελούνται τοπικά με κοινά scripts. Μόνο τα συγκεντρωτικά αποτελέσματα μοιράζονται, διασφαλίζοντας την προστασία προσωπικών δεδομένων.

Πλεονεκτήματα federated analytics:

Πίνακας 50 Πλεονεκτήματα Τεχνολογίας Federated

Πλεονέκτημα	Περιγραφή
Προστασία δεδομένων	Τα πρωτογενή δεδομένα δεν εγκαταλείπουν το ίδρυμα
Συμμόρφωση GDPR	Πλήρης συμμόρφωση με τον Κανονισμό Προστασίας Δεδομένων
Μεγαλύτερο δείγμα	Δυνατότητα ανάλυσης σε εκατομμύρια ασθενείς παγκοσμίως
Σπάνια νοσήματα	Κρίσιμο για νοσήματα με λίγους ασθενείς ανά χώρα
Τυποποίηση	Κοινό μοντέλο διασφαλίζει συγκρισιμότητα αποτελεσμάτων

6.4.4 Decentralized Analysis Workflows

Για την υλοποίηση αποκεντρωμένων αναλύσεων, το σύστημα υποστηρίζει την εκτέλεση τυποποιημένων ερωτημάτων OHDSI σε τοπικό επίπεδο. Η διαδικασία περιλαμβάνει:

1. Ορισμός ανάλυσης: Ένας κεντρικός coordinator ορίζει το ερώτημα χρησιμοποιώντας τα εργαλεία ATLAS ή HADES του OHDSI.

2. Διανομή: Το ερώτημα (SQL ή R script) διανέμεται σε όλους τους συμμετέχοντες κόμβους.
3. Τοπική εκτέλεση: Κάθε κόμβος εκτελεί το ερώτημα στη δική του OMOP βάση.
4. Συγκέντρωση: Τα συγκεντρωτικά αποτελέσματα (counts, statistics) επιστρέφονται στον coordinator.
5. Meta-analysis: Τα αποτελέσματα συνδυάζονται για την παραγωγή συνολικών συμπερασμάτων.

Αυτή η προσέγγιση είναι ιδιαίτερα σημαντική για την ανάλυση σπάνιων νοσημάτων και σπάνιων γενετικών παραλλαγών, όπου ο αριθμός ασθενών σε κάθε ίδρυμα είναι περιορισμένος αλλά η συνολική ανάλυση μπορεί να αποκαλύψει σημαντικές συσχετίσεις.

6.5 Επαναχρησιμοποιήσιμα Στοιχεία

Ένας βασικός στόχος της υλοποίησης ήταν η δημιουργία επαναχρησιμοποιήσιμων στοιχείων που διευκολύνουν την κλιμάκωση σε νέα μητρώα και νέους τύπους δεδομένων.

6.5.1 ETL Pipeline Components

Αναπτύχθηκαν modular components για τη δημιουργία ETL pipelines που μπορούν να προσαρμοστούν για διαφορετικές πηγές δεδομένων:

Πίνακας 51 ETL Components

Component	Λειτουργία	Επαναχρησιμοποίηση
LOVD Parser	Ανάγνωση δεδομένων από LOVD databases	Για άλλες LOVD βάσεις
Excel/CSV Importer	Εισαγωγή από spreadsheets	Για legacy δεδομένα
FHIR Bundle Generator	Δημιουργία FHIR resources	Για οποιοδήποτε μητρώο
OMOP ETL	Μετασχηματισμός FHIR → OMOP	Universal
HPO Resolver	Επίλυση HPO όρων	Για οποιοδήποτε phenotype data
ACMG Mapper	Αντιστοίχιση ACMG classifications	Universal για variants

6.5.2 Cohort Definition Library

Δημιουργήθηκε βιβλιοθήκη ορισμών κοορτών που μπορούν να συνδυαστούν για τη δημιουργία νέων μητρώων:

Πίνακας 52 Cohort Definition

Κατηγορία	Παραδείγματα Κανόνων
Διαγνώσεις	ICD-10 ranges (C00-D49 για neoplasms), OMIM codes
Γενετικά κριτήρια	Συγκεκριμένα γονίδια, ACMG class (P/LP), specific variants
Φαινοτυπικά	HPO terms, age of onset ranges
Δημογραφικά	Gender, age ranges, geographic origin
Temporal	Date ranges, follow-up duration

7. Κλιμάκωση και Επεκτασιμότητα

Η αρχιτεκτονική του μηχανισμού δημιουργίας μητρώων σχεδιάστηκε με κεντρικό γνώμονα την κλιμακωσιμότητα (scalability) και την επεκτασιμότητα (extensibility). Αυτές οι ιδιότητες είναι κρίσιμες για τη μακροπρόθεσμη βιωσιμότητα του συστήματος, καθώς επιτρέπουν τόσο την αύξηση του όγκου δεδομένων και χρηστών όσο και την προσθήκη νέων λειτουργιών χωρίς ανασχεδιασμό της υποδομής. Το παρόν κεφάλαιο αναλύει τις στρατηγικές κλιμάκωσης, τους μηχανισμούς επέκτασης σε νέα μητρώα, τη συμμετοχή σε federated δίκτυα, και τις μελλοντικές κατευθύνσεις ανάπτυξης.

7.1 Επέκταση σε Νέα Μητρώα

Ο μηχανισμός σχεδιάστηκε με γνώμονα την ευκολία δημιουργίας νέων μητρώων. Η διαδικασία δημιουργίας νέου μητρώου βασίζεται σε templates και επαναχρησιμοποιήσιμα components, ελαχιστοποιώντας τον χρόνο υλοποίησης και την τεχνική πολυπλοκότητα. Η modular αρχιτεκτονική επιτρέπει την προσθήκη νέων μητρώων χωρίς επίπτωση στη λειτουργία των υφιστάμενων, διασφαλίζοντας τη σταθερότητα του συστήματος.

7.1.1 Template-Based Creation

Η δημιουργία νέων μητρώων υποστηρίζεται μέσω προκαθορισμένων templates που ενσωματώνουν βέλτιστες πρακτικές για συγκεκριμένους τύπους μητρώων. Κάθε template περιλαμβάνει τον ορισμό της δομής δεδομένων (FHIR profiles, extensions), τους κανόνες επικύρωσης (validation rules), τα προκαθορισμένα value sets, τους κανόνες ETL για τον μετασχηματισμό δεδομένων, και τα dashboard templates για την απεικόνιση αποτελεσμάτων. Ο χρήστης επιλέγει ένα template που αντιστοιχεί στον τύπο μητρώου, προσαρμόζει τα κριτήρια κοόρτης και τα πεδία dataset, και ενεργοποιεί το μητρώο.

7.1.2 Self-Service για Κλινικούς Ερευνητές

Βασικός στόχος του συστήματος είναι η δυνατότητα δημιουργίας και διαχείρισης μητρώων από κλινικούς ερευνητές χωρίς εξειδικευμένη τεχνική υποστήριξη. Η γραφική διεπαφή του Cohort Builder και η βιβλιοθήκη κανόνων επιτρέπουν σε χρήστες χωρίς προγραμματιστικές γνώσεις να ορίσουν κριτήρια εισαγωγής και αποκλεισμού, να επιλέξουν τα πεδία δεδομένων προς συλλογή, να παρακολουθούν την ποιότητα δεδομένων, και να εξάγουν αναφορές και datasets για ανάλυση.

7.1.3 Διαδικασία Δημιουργίας Νέου Μητρώου

Η δημιουργία νέου μητρώου ακολουθεί μια τυποποιημένη διαδικασία πέντε βημάτων που διασφαλίζει την ποιότητα και τη συμμόρφωση με τα πρότυπα του συστήματος. Το πρώτο βήμα είναι ο Ορισμός Σκοπού και Κοόρτης, όπου καθορίζονται ο επιστημονικός σκοπός, τα κριτήρια εισαγωγής/αποκλεισμού, και ο εκτιμώμενος πληθυσμός-στόχος. Το δεύτερο βήμα αφορά την Επιλογή Template, όπου επιλέγεται το κατάλληλο template και προσαρμόζεται στις συγκεκριμένες ανάγκες. Στο τρίτο βήμα, τις Τεχνικές Προδιαγραφές, ορίζονται τα FHIR profiles, οι κανόνες validation, και τα ETL pipelines. Το τέταρτο βήμα είναι το Pilot Testing, όπου εκτελούνται δοκιμές με δείγμα δεδομένων και επικυρώνεται η λειτουργικότητα. Τέλος, στο πέμπτο βήμα, την Παραγωγική Λειτουργία, ενεργοποιείται η παραγωγική λειτουργία και ξεκινά η συλλογή δεδομένων.

Πίνακας 53 Διαδικασία Δημιουργίας Μητρώου

Φάση	Διάρκεια	Υπεύθυνοι	Παραδοτέα
1. Ορισμός Σκοπού	1-2 εβδομάδες	Κλινικός, Scientific Committee	Registry Definition Document
2. Επιλογή Template	1 εβδομάδα	Κλινικός, Data Manager	Customized Template

3. Τεχνικές Προδιαγραφές	2-4 εβδομάδες	IT Team, Data Manager	FHIR Profiles, ETL Scripts
4. Pilot Testing	2-3 εβδομάδες	Όλοι	Test Report, Validation Results
5. Παραγωγική Λειτουργία	Συνεχής	Data Manager, IT Support	Operational Registry

7.2 Συμμετοχή σε Federated Δίκτυα

Η μακροπρόθεσμη στρατηγική του συστήματος περιλαμβάνει τη συμμετοχή σε federated δίκτυα έρευνας που θα επιτρέψουν αναλύσεις σε ευρύτερο πληθυσμό, διατηρώντας παράλληλα τον τοπικό έλεγχο των δεδομένων και τη συμμόρφωση με τους κανονισμούς προστασίας προσωπικών δεδομένων. Η federated προσέγγιση είναι ιδιαίτερα σημαντική για σπάνια νοσήματα και σπάνιες γενετικές παραλλαγές, όπου ο αριθμός ασθενών σε κάθε ίδρυμα ή χώρα είναι περιορισμένος.

7.2.1 Σύνδεση με το European Health Data Space (EHDS)

Η αρχιτεκτονική του συστήματος είναι πλήρως προετοιμασμένη για σύνδεση με το European Health Data Space (EHDS), το οποίο αναμένεται να αποτελέσει τον κεντρικό μηχανισμό ανταλλαγής δεδομένων υγείας στην Ευρωπαϊκή Ένωση. Ο Κανονισμός (ΕΕ) 2025/327 για τον Ευρωπαϊκό Χώρο Δεδομένων Υγείας θέτει συγκεκριμένες απαιτήσεις τις οποίες το σύστημα καλύπτει πλήρως.

Πίνακας 54 Συμμόρφωση με Απαιτήσεις EHDS

Απαίτηση EHDS	Τεχνική Υλοποίηση	Κατάσταση
Πρότυπο ανταλλαγής δεδομένων	HL7 FHIR R4 με EHDSxF profiles	Έτοιμο
Διαλειτουργικότητα	FHIR APIs, SNOMED CT, LOINC, ICD-10	Έτοιμο
Δευτερογενής χρήση δεδομένων	OMOP CDM για analytics, ψευδωνυμοποίηση	Έτοιμο
Διασυνοριακή πρόσβαση	Federation-ready architecture, consent management	Σχεδιασμένο
Ασφάλεια και προστασία	GDPR compliance, audit trail, encryption	Έτοιμο
Patient Summary	IPS (International Patient Summary) support	Σχεδιασμένο

Σύμφωνα με τον Κανονισμό EHDS, τα κράτη μέλη υποχρεούνται να διασφαλίσουν ότι τα συστήματα ηλεκτρονικών μητρώων υγείας υποστηρίζουν τη διαλειτουργικότητα και τη δυνατότητα μεταφοράς δεδομένων μέχρι τον Μάρτιο 2029 για patient summaries και e-prescriptions, ενώ οι τεχνικές προδιαγραφές αναμένεται να οριστικοποιηθούν μέχρι τον Μάρτιο 2027.

7.2.2 Συμμετοχή στο Δίκτυο OHDSI

Η υιοθέτηση του OMOP Common Data Model επιτρέπει τη συμμετοχή στο δίκτυο OHDSI (Observational Health Data Sciences and Informatics), το οποίο αποτελεί την μεγαλύτερη διεθνή κοινότητα για την ανάλυση δεδομένων υγείας πραγματικού κόσμου. Το δίκτυο OHDSI περιλαμβάνει περισσότερα από 300 ιδρύματα σε 30+ χώρες, με συνολικά δεδομένα από περισσότερους από 1 δισεκατομμύριο ασθενείς παγκοσμίως.

Πίνακας 55 Εργαλεία OHDSI για Federated Analytics

Εργαλείο	Λειτουργία	Χρήση στο Σύστημα
ATLAS	Web interface για ορισμό κοορτών και αναλύσεων	Ορισμός cohort definitions, characterization
HADES	R packages για στατιστική ανάλυση	Population-level estimation, prediction

Data Quality Dashboard	Αξιολόγηση ποιότητας OMOP database	Περιοδικοί έλεγχοι ποιότητας δεδομένων
ACHILLES	Automated characterization of health	Baseline statistics, data profiling
WhiteRabbit	ETL specification and mapping	Σχεδιασμός FHIR→OMOP ETL

7.2.3 Cross-Border Genomic Analytics

Η τυποποίηση μέσω OMOP και της επέκτασης Genomic CDM ανοίγει τη δυνατότητα για διασυνοριακές αναλύσεις γενετικών δεδομένων. Αυτό είναι ιδιαίτερα σημαντικό για σπάνια νοσήματα όπου ο αριθμός ασθενών σε μία χώρα είναι περιορισμένος. Η συμμετοχή σε διεθνή δίκτυα όπως τα European Reference Networks (ERNs) και το Beyond 1 Million Genomes (B1MG) initiative επιτρέπει την πρόσβαση σε μεγαλύτερα δείγματα για στατιστικά ισχυρές αναλύσεις.

Η federated προσέγγιση για γενετικά δεδομένα παρουσιάζει ιδιαίτερες προκλήσεις λόγω της ευαισθησίας των δεδομένων. Το σύστημα υλοποιεί διάφορες τεχνικές για την προστασία της ιδιωτικότητας κατά τη διεξαγωγή federated αναλύσεων: εκτέλεση των αναλύσεων τοπικά με μεταφορά μόνο συγκεντρωτικών αποτελεσμάτων, εφαρμογή minimum cell size thresholds για αποφυγή re-identification, χρήση διαφορικής ιδιωτικότητας (differential privacy) σε ορισμένες αναλύσεις, και ψευδωνυμοποίηση δεδομένων πριν τη συμμετοχή σε federated queries.

7.2.4 Αρχιτεκτονική Federation Node

Κάθε συμμετέχον ίδρυμα στο federated δίκτυο λειτουργεί ως ανεξάρτητος κόμβος (node) που διατηρεί πλήρη έλεγχο των δεδομένων του. Η αρχιτεκτονική του node περιλαμβάνει τα εξής επίπεδα: το Data Layer που περιλαμβάνει τον τοπικό FHIR Server για επιχειρησιακά δεδομένα και την OMOP database για analytics, το Processing Layer που αναλαμβάνει την τοπική εκτέλεση standardized scripts (R, SQL) και τους ελέγχους ποιότητας δεδομένων, το Security Layer με τη διαχείριση πρόσβασης, τους ελέγχους συγκατάθεσης, και την κρυπτογράφηση, και τέλος το Communication Layer που χειρίζεται τα secure APIs για επικοινωνία με τον central coordinator και την ανταλλαγή αποτελεσμάτων με άλλους κόμβους.

7.3 Τεχνική Κλιμάκωση

Η τεχνική αρχιτεκτονική του συστήματος σχεδιάστηκε για να υποστηρίζει την αύξηση του όγκου δεδομένων και του αριθμού χρηστών χωρίς υποβάθμιση της απόδοσης. Οι βασικές στρατηγικές κλιμάκωσης περιλαμβάνουν τόσο οριζόντια όσο και κάθετη επέκταση.

7.3.1 Στρατηγικές Κλιμάκωσης

Πίνακας 56 Στρατηγικές Κλιμάκωσης

Στρατηγική	Περιγραφή	Εφαρμογή
Horizontal Scaling	Προσθήκη περισσότερων instances για κατανομή φορτίου	FHIR Server cluster, load balancing
Vertical Scaling	Αύξηση πόρων (CPU, RAM) σε υπάρχοντα servers	Database server upgrades
Database Sharding	Κατανομή δεδομένων σε πολλαπλές βάσεις	Partitioning by registry, by date range
Caching	Αποθήκευση συχνά χρησιμοποιούμενων δεδομένων	Redis για session data, query results
Asynchronous Processing	Εκτέλεση βαρέων εργασιών στο background	ETL jobs, report generation

CDN Integration	Κατανεμημένη παράδοση στατικού περιεχομένου	Dashboard assets, documentation
-----------------	---	---------------------------------

7.3.2 Εκτιμώμενη Χωρητικότητα

Η τρέχουσα υποδομή έχει σχεδιαστεί για να υποστηρίζει τις ανάγκες του έργου, με δυνατότητα κλιμάκωσης για μελλοντική επέκταση. Οι εκτιμήσεις χωρητικότητας βασίζονται σε πραγματικές μετρήσεις απόδοσης και διεθνείς benchmarks για παρόμοια συστήματα.

Πίνακας 57 Εκτιμήσεις Χωρητικότητας Συστήματος

Μετρική	Τρέχουσα Χωρητικότητα	Με Κλιμάκωση
Αριθμός ασθενών	~10.000	~100.000
Γενετικές παραλλαγές	~50.000	~500.000
Concurrent users	~50	~500
API requests/sec	~100	~1.000
Storage (FHIR)	~50 GB	~500 GB
Storage (OMOP)	~20 GB	~200 GB

7.4 Μελλοντικές Επεκτάσεις

Πέρα από τις υλοποιημένες και προγραμματισμένες λειτουργίες, η αρχιτεκτονική επιτρέπει μια σειρά μελλοντικών επεκτάσεων που θα ενισχύσουν περαιτέρω τις δυνατότητες του συστήματος. Οι επεκτάσεις αυτές κατηγοριοποιούνται σε τρεις χρονικούς ορίζοντες.

7.4.1 Βραχυπρόθεσμες Επεκτάσεις

Οι βραχυπρόθεσμες επεκτάσεις επικεντρώνονται στην ολοκλήρωση της βασικής λειτουργικότητας και τη βελτίωση της εμπειρίας χρήστη. Περιλαμβάνουν την ολοκλήρωση του OMOP ETL pipeline με πλήρη υποστήριξη Genomic CDM, την ανάπτυξη self-service Cohort Builder UI για κλινικούς, την ενσωμάτωση automated data quality dashboards, και την υλοποίηση notification system για κρίσιμα ευρήματα.

7.4.2 Μεσοπρόθεσμες Επεκτάσεις

Οι μεσοπρόθεσμες επεκτάσεις στοχεύουν στην ενίσχυση των αναλυτικών δυνατοτήτων και τη διεύρυνση του πεδίου εφαρμογής. Η πρώτη επέκταση αφορά την ενσωμάτωση AI/ML μοντέλων για αυτόματη ταξινόμηση παραλλαγών αβέβαιης κλινικής σημασίας (VUS reclassification) και πρόβλεψη κλινικών εκβάσεων με βάση γενετικά και φαινοτυπικά δεδομένα. Η δεύτερη επέκταση περιλαμβάνει την υποστήριξη επιπλέον τύπων ομικών δεδομένων, συμπεριλαμβανομένων μεταβολομικών δεδομένων για μεταβολικά νοσήματα, πρωτεομικών δεδομένων για ογκολογία, και transcriptomics για RNA-based διαγνωστικά. Η τρίτη επέκταση αφορά τη διασύνδεση με βιοτράπεζες, επιτρέποντας τη σύνδεση κλινικών δεδομένων με βιολογικά δείγματα και υποστήριξη sample request workflows.

7.4.3 Μακροπρόθεσμες Επεκτάσεις

Οι μακροπρόθεσμες επεκτάσεις στοχεύουν στην πλήρη ενσωμάτωση σε ευρωπαϊκές υποδομές και την ανάπτυξη προηγμένων δυνατοτήτων. Αυτές περιλαμβάνουν την πλήρη ενσωμάτωση στο EHDS ως Health Data Access Body (HDAB) node, την ανάπτυξη real-time CDS (Clinical Decision Support) με ενσωμάτωση στα νοσοκομειακά πληροφοριακά συστήματα, και την υλοποίηση advanced analytics platform με federated machine learning capabilities.

7.5 Σύνοψη

Το σύστημα μητρώων που αναπτύχθηκε στα πλαίσια του έργου TAEDR-0539180 παρέχει μια στιβαρή, κλιμακώσιμη και επεκτάσιμη υποδομή για την Ιατρική Ακριβείας στην Ελλάδα. Η αρχιτεκτονική διπλής στρώσης (FHIR/OMOP), η υιοθέτηση διεθνών προτύπων, και η σχεδίαση για federation διασφαλίζουν τη μακροπρόθεσμη βιωσιμότητα και τη δυνατότητα ενσωμάτωσης σε ευρωπαϊκές υποδομές.

Τα βασικά πλεονεκτήματα της αρχιτεκτονικής περιλαμβάνουν την ευκολία δημιουργίας νέων μητρώων μέσω templates και self-service εργαλείων, την ετοιμότητα για συμμετοχή σε federated δίκτυα όπως το OHDSI και το EHDS, την τεχνική κλιμακωσιμότητα για υποστήριξη αυξανόμενου όγκου δεδομένων και χρηστών, και τη σαφή στρατηγική για μελλοντικές επεκτάσεις που θα ενισχύσουν περαιτέρω τις δυνατότητες του συστήματος. Η επιτυχής υλοποίηση των πιλοτικών μητρώων WES και CanVaS αποδεικνύει τη λειτουργικότητα της υποδομής και θέτει τις βάσεις για την επέκταση σε νέους τομείς της Ιατρικής Ακριβείας.

8. Συμπεράσματα

8.1 Σύνοψη Επιτευγμάτων

Το παραδοτέο Π12.2 ολοκληρώνει επιτυχώς την ανάπτυξη ενός ευφυούς μηχανισμού δημιουργίας μητρώων ασθενών, ο οποίος λειτουργεί σε πλήρη διαλειτουργικότητα με τον Έξυπνο Ηλεκτρονικό Φάκελο Υγείας (Π12.1). Η υλοποίηση υπερβαίνει τις αρχικές προδιαγραφές, παρέχοντας μια ολοκληρωμένη πλατφόρμα για τη δημιουργία, διαχείριση και ανάλυση μητρώων με γενετικά δεδομένα.

Τα κύρια επιτεύγματα του παραδοτέου περιλαμβάνουν την υλοποίηση αρχιτεκτονικής διπλής στρώσης που συνδυάζει το πρότυπο HL7 FHIR R4 για την κλινική λειτουργία με το OMOP Common Data Model για αναλύσεις δευτερογενούς χρήσης, την ενσωμάτωση της επέκτασης Genomic CDM του OHDSI για την τυποποιημένη αποθήκευση γενετικών δεδομένων, την ανάπτυξη δύο πιλοτικών μητρώων με συνολικά περισσότερους από 7.800 ασθενείς, τη δημιουργία επαναχρησιμοποιήσιμων ETL components για κλιμάκωση σε νέα μητρώα, καθώς και την πλήρη συμμόρφωση με τον Κανονισμό GDPR και ετοιμότητα για ενσωμάτωση στον Ευρωπαϊκό Χώρο Δεδομένων Υγείας (EHDS).

Το WES Registry περιλαμβάνει περίπου 300 ασθενείς με σπάνια νοσήματα που αναλύθηκαν με αλληλούχιση ολόκληρου του εξωνιώματος, ενώ το CanVaS Registry ενσωματώνει γενετικά δεδομένα από 7.363 Έλληνες με προδιάθεση σε καρκίνο, καλύπτοντας 97 γονίδια και περίπου 24.000 καταχωρήσεις παραλλαγών. Η επιτυχής υλοποίηση αυτών των μητρώων αποδεικνύει τη λειτουργικότητα του μηχανισμού και θέτει τις βάσεις για μελλοντική επέκταση.

8.2 Καινοτομία

Ο μηχανισμός που αναπτύχθηκε παρουσιάζει σημαντική καινοτομία τόσο σε τεχνολογικό όσο και σε επιστημονικό επίπεδο. Η αρχιτεκτονική διπλής στρώσης αποτελεί πρωτοποριακή προσέγγιση που επιτρέπει την ταυτόχρονη εξυπηρέτηση κλινικών και ερευνητικών αναγκών από μια ενιαία υποδομή, διατηρώντας παράλληλα τη βέλτιστη απόδοση για κάθε περίπτωση χρήσης.

Η υιοθέτηση του Genomic CDM τοποθετεί το σύστημα στην πρώτη γραμμή της διεθνούς τυποποίησης γενωμικών δεδομένων υγείας. Η επέκταση αυτή, που αναπτύχθηκε από την κοινότητα OHDSI, επιτρέπει τη συμμετοχή σε federated δίκτυα έρευνας και τη διεξαγωγή αποκεντρωμένων αναλύσεων χωρίς μεταφορά πρωτογενών δεδομένων. Αυτό είναι ιδιαίτερα σημαντικό για γενετικά δεδομένα, όπου η προστασία της ιδιωτικότητας αποτελεί θεμελιώδη απαίτηση.

Επιπλέον, η ανάπτυξη ειδικών FHIR extensions για τον ελληνικό πληθυσμό, συμπεριλαμβανομένης της καταγραφής founder mutations και της γεωγραφικής κατανομής παραλλαγών, αποτελεί σημαντική συμβολή στην Ιατρική Ακριβείας σε εθνικό επίπεδο. Αυτές οι επεκτάσεις επιτρέπουν την αξιοποίηση πληθυσμιακά-ειδικών δεδομένων που δεν είναι διαθέσιμα σε διεθνείς βάσεις όπως το gnomAD ή το ClinVar.

8.3 Συμβολή στους Στόχους του Έργου

Το παραδοτέο συμβάλλει ουσιαστικά στην επίτευξη των κεντρικών στόχων του έργου TAEDR-0539180. Ο Στόχος 4, που αφορά την ενσωμάτωση δεδομένων σε ενιαίο οικοσύστημα μητρώων νοσημάτων, υλοποιείται πλήρως μέσω του μηχανισμού δημιουργίας μητρώων και των δύο πιλοτικών εφαρμογών. Η αρχιτεκτονική διασφαλίζει τη διαλειτουργικότητα μεταξύ μητρώων και την κοινή χρήση δεδομένων σε τυποποιημένη μορφή.

Ο Στόχος 5 για την ανάπτυξη Έξυπνου Ηλεκτρονικού Φακέλου Υγείας με ενσωματωμένες λειτουργίες δευτερογενούς ανάλυσης επιτυγχάνεται μέσω της διασύνδεσης με το Π12.1 και των αναλυτικών δυνατοτήτων που παρέχει το επίπεδο OMOP. Η υποστήριξη για Health Technology Assessment αποτελεί βασικό στοιχείο αυτής της επίτευξης, επιτρέποντας την αξιολόγηση τεχνολογιών υγείας με δεδομένα πραγματικής κλινικής πρακτικής.

Τέλος, ο Στόχος 6 για τη δημιουργία διεπιστημονικού Hub ανάλυσης δεδομένων Ιατρικής Ακριβείας υποστηρίζεται από τη δυνατότητα συνεργασίας μέσω federated analytics και την ετοιμότητα για

ενσωμάτωση σε ευρωπαϊκά δίκτυα. Η συμμετοχή στα δίκτυα OHDSI και EHDS θα επεκτείνει τις δυνατότητες ανάλυσης σε εκατομμύρια ασθενείς παγκοσμίως.

8.4 Μελλοντική Ανάπτυξη

Η ολοκλήρωση του παραδοτέου θέτει στέρεες βάσεις για μελλοντική ανάπτυξη σε πολλαπλές κατευθύνσεις. Βραχυπρόθεσμα, προγραμματίζεται η επέκταση σε περισσότερα μητρώα που θα καλύπτουν επιπλέον νοσήματα και πληθυσμούς, αξιοποιώντας τα επαναχρησιμοποιήσιμα components και templates που αναπτύχθηκαν. Ιδιαίτερη έμφαση θα δοθεί σε μητρώα για καρδιαγγειακά νοσήματα και μεταβολικά σύνδρομα, που αποτελούν κεντρικούς άξονες του έργου.

Μεσοπρόθεσμα, η σύνδεση με τον Ευρωπαϊκό Χώρο Δεδομένων Υγείας (EHDS) θα αποτελέσει κεντρική προτεραιότητα. Σύμφωνα με τον Κανονισμό (ΕΕ) 2025/327, τα κράτη μέλη καλούνται να διασφαλίσουν τη διαλειτουργικότητα των συστημάτων ηλεκτρονικών μητρώων υγείας μέχρι τον Μάρτιο 2029. Η αρχιτεκτονική που αναπτύχθηκε πληροί ήδη τις βασικές απαιτήσεις, τοποθετώντας το σύστημα σε πλεονεκτική θέση για την έγκαιρη ενσωμάτωση.

Μακροπρόθεσμα, η ενσωμάτωση τεχνικών τεχνητής νοημοσύνης και μηχανικής μάθησης θα ενισχύσει περαιτέρω τις αναλυτικές δυνατότητες του συστήματος. Η αυτόματη επαναταξινόμηση παραλλαγών αβέβαιης σημασίας (VUS reclassification), η πρόβλεψη κλινικών εκβάσεων, και η ανάπτυξη εργαλείων υποστήριξης κλινικής απόφασης αποτελούν φυσικές επεκτάσεις που θα μεγιστοποιήσουν την αξία των δεδομένων που συλλέγονται. Παράλληλα, η δημιουργία ολοκληρωμένων εργαλείων HTA θα υποστηρίξει τη λήψη τεκμηριωμένων αποφάσεων πολιτικής υγείας σε εθνικό επίπεδο.

Συνολικά, ο μηχανισμός δημιουργίας μητρώων που αναπτύχθηκε αποτελεί θεμέλιο λίθο για την εφαρμογή της Ιατρικής Ακριβείας στην Ελλάδα. Η συνδυαστική αξιοποίηση κλινικών και γενετικών δεδομένων μέσω τυποποιημένων προτύπων διασφαλίζει τη βιωσιμότητα και την κλιμακωσιμότητα του συστήματος, ενώ η ετοιμότητα για συμμετοχή σε ευρωπαϊκές υποδομές ανοίγει νέες προοπτικές για έρευνα και κλινική πρακτική.

9. Αναφορές

9.1 Πρότυπα και Specifications

1. HL7 International. HL7 FHIR R4 Specification. Release 4.0.1. 2019. Available from: <https://hl7.org/fhir/R4/>
2. OHDSI. OMOP Common Data Model. Version 5.4. 2023. Available from: <https://ohdsi.github.io/CommonDataModel/>
3. OHDSI Genomics Working Group. Genomic CDM Extension. 2023. Available from: <https://github.com/OHDSI/Genomic-CDM>
4. European Parliament and Council. Regulation (EU) 2025/327 on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2023/2854. Official Journal of the European Union. 2025.
5. European Parliament and Council. Regulation (EU) 2016/679 (General Data Protection Regulation - GDPR). Official Journal of the European Union. 2016.
6. HL7 International. Genomics Reporting Implementation Guide (STU2). Release 2.0.0. 2022. Available from: <https://hl7.org/fhir/uv/genomics-reporting/STU2/>
7. European Commission. European Health Data Space - Technical Specifications for Electronic Health Record Systems. 2024. Available from: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space-regulation-ehds_en

9.2 Μητρώα και Δίκτυα

8. Gliklich RE, Dreyer NA, Leavy MB, editors. Registries for Evaluating Patient Outcomes: A User's Guide [Internet]. 3rd edition. Rockville (MD): Agency for Healthcare Research and Quality (AHRQ); 2014 Apr. AHRQ Publication No. 13(14)-EHC111. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK208616/>
9. OHDSI. Observational Health Data Sciences and Informatics. Available from: <https://ohdsi.org/>
10. Hripesak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. Stud Health Technol Inform. 2015;216:574-578. doi:10.3233/978-1-61499-564-7-574
11. OHDSI Collaborative. The Book of OHDSI. 2021. Available from: <https://ohdsi.github.io/TheBookOfOhdsi/>
12. European Commission. PARENT Joint Action - Cross-border PATient REGistries iNiTiative. 2015. Available from: <https://eu-rd-platform.jrc.ec.europa.eu/>
13. European Commission. European Reference Networks (ERNs). Available from: https://health.ec.europa.eu/rare-diseases-and-european-reference-networks/european-reference-networks_en
14. National Cancer Institute. SEER Program - Surveillance, Epidemiology, and End Results. NIH. Available from: <https://seer.cancer.gov/>

15. INSERM. Orphanet - The portal for rare diseases and orphan drugs. Available from: <https://www.orpha.net/>

9.3 Coding Systems

16. Regenstrief Institute. LOINC - Logical Observation Identifiers Names and Codes. Available from: <https://loinc.org/>
17. SNOMED International. SNOMED CT. Available from: <https://www.snomed.org/>
18. Köhler S, Gargano M, Matentzoglou N, et al. The Human Phenotype Ontology in 2021. *Nucleic Acids Res.* 2021;49(D1):D1207-D1217. doi:10.1093/nar/gkaa1043
19. Johns Hopkins University. OMIM - Online Mendelian Inheritance in Man. Available from: <https://www.omim.org/>
20. World Health Organization. ICD-10 - International Classification of Diseases, 10th Revision. Available from: <https://icd.who.int/>
21. den Dunnen JT, Dalgleish R, Maglott DR, et al. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum Mutat.* 2016;37(6):564-569. doi:10.1002/humu.22981

9.4 Γενετική και Ταξινόμηση Παραλλαγών

22. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17(5):405-424. doi:10.1038/gim.2015.30
23. Li MM, Datto M, Duncavage EJ, et al. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *J Mol Diagn.* 2017;19(1):4-23. doi:10.1016/j.jmoldx.2016.10.002
24. Tavtigian SV, Greenblatt MS, Harrison SM, et al. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet Med.* 2018;20(9):1054-1060. doi:10.1038/gim.2017.210
25. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* 2018;46(D1):D1062-D1067. doi:10.1093/nar/gkx1153

Παράρτημα Α: Γλωσσάρι

Όρος	Ορισμός
CDM	Common Data Model - Κοινό μοντέλο δεδομένων
Cohort	Κοόρτη - Ομάδα ασθενών με κοινά χαρακτηριστικά
EHDS	European Health Data Space - Ευρωπαϊκός Χώρος Δεδομένων Υγείας
ETL	Extract-Transform-Load - Διαδικασία εξαγωγής, μετασχηματισμού και φόρτωσης δεδομένων
Federated	Κατανεμημένη ανάλυση χωρίς μεταφορά δεδομένων
FHIR	Fast Healthcare Interoperability Resources - Πρότυπο διαλειτουργικότητας
Founder Mutation	Μετάλλαξη με υψηλή συχνότητα σε συγκεκριμένο πληθυσμό λόγω κοινής καταγωγής
GDPR	General Data Protection Regulation - Γενικός Κανονισμός Προστασίας Δεδομένων
HTA	Health Technology Assessment - Αξιολόγηση Τεχνολογιών Υγείας
KPI	Key Performance Indicator - Βασικός δείκτης απόδοσης
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership - Common Data Model
Provenance	Προέλευση δεδομένων - Πληροφορίες για την πηγή και ιστορικό δεδομένων
RBAC	Role-Based Access Control - Έλεγχος πρόσβασης βάσει ρόλων
Registry	Μητρώο - Οργανωμένο σύστημα συλλογής δεδομένων υγείας
RWD	Real World Data - Δεδομένα πραγματικής κλινικής πρακτικής
RWE	Real World Evidence - Τεκμήρια από δεδομένα πραγματικού κόσμου
WES	Whole Exome Sequencing - Αλληλούχιση εξονίων

Παράρτημα Β: Πίνακες Genomic CDM

B.1 TARGET_GENE

Ο πίνακας TARGET_GENE περιέχει πληροφορίες για τα γονίδια που εξετάστηκαν σε κάθε ασθενή.

Πεδίο	Τύπος	Περιγραφή
target_gene_id	INTEGER	Primary key
person_id	INTEGER	Foreign key to PERSON
gene_concept_id	INTEGER	Concept ID for gene (HGNC)
gene_source_value	VARCHAR	Gene symbol (e.g., BRCA1)
test_type_concept_id	INTEGER	Type of test (WES, WGS, panel)
result_concept_id	INTEGER	Result (positive, negative, VUS)
test_date	DATE	Date of testing

B.2 VARIANT_OCCURRENCE

Ο πίνακας VARIANT_OCCURRENCE καταγράφει τις γενετικές παραλλαγές που εντοπίστηκαν.

Πεδίο	Τύπος	Περιγραφή
variant_occurrence_id	INTEGER	Primary key
person_id	INTEGER	Foreign key to PERSON
target_gene_id	INTEGER	Foreign key to TARGET_GENE
reference_sequence	VARCHAR	Reference sequence (e.g., NM_007294.4)
hgvs_c	VARCHAR	HGVS DNA notation (e.g., c.5266dupC)
hgvs_p	VARCHAR	HGVS protein notation (e.g., p.Gln1756fs)
zygosity_concept_id	INTEGER	Zygosity (heterozygous, homozygous)
allele_frequency	FLOAT	Population allele frequency

B.3 VARIANT_ANNOTATION

Ο πίνακας VARIANT_ANNOTATION περιέχει τις ερμηνείες και ταξινομήσεις των παραλλαγών.

Πεδίο	Τύπος	Περιγραφή
variant_annotation_id	INTEGER	Primary key
variant_occurrence_id	INTEGER	Foreign key to VARIANT_OCCURRENCE
classification_concept_id	INTEGER	ACMG classification
classification_source	VARCHAR	Source (ClinVar, laboratory)
evidence_codes	VARCHAR	ACMG evidence codes
annotation_date	DATE	Date of annotation