


CASEIB
2022

XL Congreso
Anual de la
Sociedad
Española de
Ingeniería
Biomédica

40 AÑOS UNIENDO FUERZAS
PARA IMPULSAR LA
INGENIERÍA BIOMÉDICA

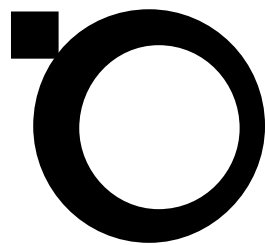
23-25 NOVIEMBRE 2022 | VALLADOLID

LIBRO DE ACTAS



GRUPO DE
INGENIERÍA
BIOMÉDICA
UNIVERSIDAD DE VALLADOLID





**CASEIB
2022**

**XL Congreso Anual
de la Sociedad
Española de
Ingeniería
Biomédica**

23-25 NOVIEMBRE 2022 | VALLADOLID

XL Congreso Anual de la Sociedad Española de Ingeniería Biomédica CASEIB 2022

Los trabajos incluidos en esta publicación han sido evaluados por el Comité Científico que en ella se relaciona y según el procedimiento que se recoge en www.caseib.es/2022/

Presidente del Comité Organizador:

Roberto Hornero Sánchez

Secretarios del Comité Organizador:

Jesús Poza Crespo

Carlos Gómez Peña

Editores:

Roberto Hornero Sánchez

Jesús Poza Crespo

Carlos Gómez Peña

María García Gadañón

Maquetación:

Aarón Maturana Candelas

Carlos Gómez Peña

Diseño de portada:

Sandra Martínez Cagigal

Víctor Martínez Cagigal

ISBN: 978-84-09-45972-8



Libro de Actas del XL Congreso Anual de la Sociedad Española de Ingeniería Biomédica

Se distribuye bajo una licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional.

Transformación OMOP-CDM del repositorio para usos secundarios de la cohorte de la Red Española de Investigación en SIDA (CoRIS)

M. Pascual-Carrasco^{1,4}, A. Burgos Llamo¹, I. Jarrín Vera^{2,4}, M. Pedrera-Jiménez³, N. García-Barrio³, R. Sánchez-de Madariaga¹, A. Muñoz Carrero^{1,4}

¹ Unidad de Investigación en Salud Digital. Instituto de Salud Carlos III, Madrid, España, mario.pascual@isciii.es

² Centro Nacional de Epidemiología, Instituto de Salud Carlos III, Madrid, España

³ Unidad de Ciencia de Datos, Hospital Universitario 12 de octubre, Madrid, España

⁴ CIBER de Enfermedades Infecciosas, CIBERINFEC, Madrid, España

Resumen

CoRIS es el registro longitudinal de datos clínico-epidemiológicos de la cohorte española de investigación SIDA. CoRIS es conforme a HICDEP, modelo relacional mono-propósito no alineado con las características que constituyen las bases esenciales en repositorios avanzados para usos secundarios: capacidad de integración flexible de nuevas fuentes de datos, interoperabilidad semántica y diseño multi-propósito. Este aspecto supone una dificultad para la evolución de CoRIS en relación con la reutilización y uso compartido de datos y la replicación de estudios. En el ámbito de los usos secundarios, se posicionan como referencia los modelos comunes de datos (CDM); entre ellos, destaca OMOP-CDM. A partir de CoRIS/HICDEP, se ha implementado una nueva instancia conforme al modelo OMOP-CDM. Se ha evaluado la nueva instancia CoRIS/OMOP frente a CoRIS/HICDEP y se ha demostrado no-inferioridad en relación con: capacidad de cobertura de conceptos; integridad de datos; y capacidad para replicar cohortes complejas. De 410 conceptos clínico-epidemiológicos en CoRIS, se ha alcanzado una cobertura semántica óptima en un 79% y sub-óptima 20,2%. Se ha verificado la integridad de los datos al 100% (identificación y extracción). Se ha replicado de forma precisa las cohortes de tres estudios científicos publicados en dominios combinados relevantes para CoRIS (análisis de enfermedades defintorias SIDA; pautas antirretrovirales; pruebas de laboratorio). CoRIS/OMOP demuestra ser un repositorio no-inferior a CoRIS/HICDEP, adicionalmente aportando a CoRIS capacidad semántica en persistencia y explotación (reutilización y replicabilidad de evidencia), y escalabilidad en volumen e integración de nuevas fuentes y dominios de datos sin modificaciones en su modelo de datos.

1. Introducción

Desde organismos e instituciones internacionales, se promueve la persistencia e integración de los datos en repositorios sujetos a los principios FAIR para compartir, reutilizar, y conformar conjuntos de datos fiables, precisos, normalizados, anonimizados, accesibles, semánticamente interoperables y puestos a disposición de la comunidad científica para la generación y replicación de nueva evidencia, la creación de redes de conocimiento biomédico, dirigidas a maximizar el impacto y los beneficios de la investigación en las necesidades de salud de la sociedad [1].

En el ámbito de la persistencia de los datos para usos secundarios, instituciones de investigación, sanitarias, consorcios científicos, despliegan repositorios y redes de repositorios conformes a modelos de datos comunes (CDM), multi-propósito, que posibilitan el almacenamiento e intercambio de datos, escalado y explotación en entornos confiables, y con alto potencial colaborativo [2,3]. OMOP (Observational Medical Outcomes Partnership), impulsado por OHDSI (Observational Health Data Sciences and Informatics), se sitúa entre los CDM semánticamente interoperables para persistencia y explotación en usos secundarios en registros longitudinales de datos de salud [4].

CoRIS [5] es una cohorte abierta, multicéntrica y prospectiva de ámbito nacional constituida por sujetos reclutados en unidades asistenciales de VIH del Sistema Sanitario Público español, con infección confirmada por el VIH y sin tratamiento antirretroviral en el momento de la inclusión. Desde 2004, CoRIS acumula información longitudinal de datos clínico-epidemiológicos de más de 16 mil pacientes procedentes de 45 hospitales en 13 CCAA, constituyendo un recurso fundamental en la investigación clínico-epidemiológica VIH/ SIDA.

El repositorio de datos CoRIS es conforme a HICDEP (HIV Cohorts Data Exchange Protocol) [6]. HICDEP es un modelo de persistencia e intercambio de datos consensuado entre cohortes europeas de investigación, con desarrollo desde 2011, que responde a una aproximación tradicional de diseño relacional mono-propósito (evolución sujeta a cambios en el propio modelo), y mínimo soporte semántico normalizado (codificaciones propietarias). Estos factores constituyen barreras para avanzar hacia las necesidades en relación con la gestión avanzada de datos para usos secundarios [3]: asimilación de nuevos dominios, conceptos y relaciones, reutilización, uso compartido de datos y capacidad de replicación de evidencia.

2. Objetivo

A partir del actual repositorio CoRIS basado en el modelo HICDEP (CoRIS/HICDEP), se plantea llevar a cabo:

- La transformación de la instancia CoRIS/HICDEP a CoRIS/OMOP

- Evaluación comparada CoRIS/OMOP frente a CoRIS/HICDEP en términos de no inferioridad

3. Material

La instancia CoRIS/HICDEP facilitada desde CoRIS es la versión de noviembre 2018 (CoRIS/HICDEP v2018) con datos de 15.509 pacientes, sobre ficheros DTA (STATA). Para el soporte de servicios y herramientas se ha habilitado una infraestructura telemática sobre máquinas virtuales (VMWare), S.O. Redhat Linux. Las instancias CoRIS/HICDEP y CoRIS/OMOP han sido desplegadas sobre el gestor de bases de datos relacional PostgreSQL 9.2. La infraestructura OMOP-CDM (v5.3.1) se ha puesto en marcha utilizando las herramientas públicas OHDSI [3]: creación de base de datos (tablas, índices, restricciones, carga de vocabularios); acceso a instancia OMOP: atlas (v2.8) y webAPI (v2.9); examen de calidad de la instancia OMOP: librerías R achilles/achilles heel (v1.6.7) y DQD 1.0; mapeo de conceptos con vocabularios estándar: web Athena y Usagi (v1.4.3), y otras web de terminología médica (SNOMED-CT, RXNORM, ATC, LOINC, principalmente); diseño del proceso ETL: Rabbit-in-a-hat; implementación ETL: java 1.8 (eclipse). Para la explotación de datos de la instancia OMOP se ha empleado SQL (pgAdmin 5) y R v4.0.4 (RStudio 1.4).

El equipo de investigación ha estado compuesto por médicos (dominio sanitario); ciencia e ingeniería de datos y biomédica (dominio de los datos); e informática (dominio de ingeniería de software/ sistemas).

4. Métodos

4.1. Transformación a CoRIS/OMOP

La metodología desarrollada para la transformación de CoRIS/HICDEP a CoRIS/OMOP consiste, básicamente, en un proceso ETL en 4 fases: identificación de conceptos clínico-epidemiológicos en origen y mapeo a conceptos basados en terminologías estándar; diseño del proceso ETL; implementación del proceso ETL; y evaluación de calidad de la instancia OMOP-CDM.

Para disponer de una infraestructura de persistencia homogénea, inicialmente se llevó a cabo una carga desde CoRIS/HICDEP sobre STATA a CoRIS/HICDEP sobre PostgreSQL, manteniendo la misma estructura de tablas, campos, tipos y codificaciones (extracción CSV y carga PostgreSQL); para ello fue utilizado el framework Pentaho Data Integration (Kettle).

Analizando la documentación CoRIS y la referencia HICDEP [6], se realizó la identificación y catalogación de conceptos clínico-epidemiológicos CoRIS. A partir la lista de conceptos, se llevó a cabo un proceso sistemático de mapeo semántico hacia conceptos estándar OMOP-CDM.

A continuación, se realizó el diseño del proceso ETL CoRIS/HICDEP a OMOP-CDM. Los registros OMOP-CDM destino deben replicar íntegramente los registros CoRIS/HICDEP de origen con transformaciones soportadas por la semántica y los mecanismos implícitos del CDM. Este hecho convierte el diseño ETL en un proceso que requiere diversos ciclos de iteración. Una vez

concluido el diseño, el proceso ETL ha sido implementado mediante un programa java “ad-hoc”.

Finalmente, se llevó a cabo un examen de calidad de la nueva instancia CoRIS/OMOP. Los parámetros de calidad se refieren a la completitud, conformidad y plausibilidad de los registros CoRIS/OMOP [8]. A través un extenso conjunto de sentencias SQL y scripts R (proporcionados desde OHDSI), son analizados, principalmente, aspectos sintácticos y semánticos en la nueva instancia en relación con el CDM y los vocabularios estándar. Los informes de calidad guían el proceso iterativo de depuración y mejora progresiva de la nueva instancia.

4.2. Evaluación comparada

Se estableció como hipótesis de partida la no inferioridad en capacidades de persistencia y explotación de la nueva instancia CoRIS/OMOP frente a CoRIS/HICDEP. Para corroborar la hipótesis, se planteó tomar como referencias las metodologías y marcos de trabajo propuestos por Garza [7] (enfoque genérico), Voss [8] y Vaclav [9] (enfoque a OMOP). Garza [2] establece un conjunto de categorías, criterios y métricas para llevar a cabo caracterizaciones globales de capacidades de modelos. Fueron seleccionados los siguientes criterios de capacidad: cobertura de conceptos y datos (“content coverage”); implementación de condicionantes y restricciones (“integrity”); y extracción de datos (“queriability”); se establecieron métricas orientadas a OMOP descritas a continuación.

Para determinar la cobertura de conceptos, el grado de correspondencia entre conceptos originales CoRIS/HICDEP y conceptos mapeados en CoRIS/OMOP fue cuantificado en términos de mapeo óptimo, sub-óptimo y no-mapeado. Para verificar la cobertura de datos y capacidad de extracción de datos, se tomó como patrón de referencia el informe anual emitido por CoRIS (elaborado con CoRIS/HICDEP v2018) en el que son exhaustivamente cuantificados los diferentes tipos de datos y agrupaciones de datos en la instancia: demográficos (género, edad, origen, nivel educativo, etc.); modos de infección; laboratorio (CD4, CD8, carga viral, albúmina, ALT/AST, triglicéridos, HDL, LDL, creatinina, plaquetas); virología (anticuerpos VHB, DNA de VHB, antígenos HBe,VHB, anticuerpos VHC, HCV-RNA, hepatitis delta, HIV-1, HIV western blot, sífilis RPR-VDRL); enfermedades defintorias SIDA; pautas TAR; eventos no-SIDA; y causas de muerte. Se desarrolló un conjunto de 25 sentencias SQL y scripts R para la reproducción del informe CoRIS v2018. Adicionalmente, se realizó el análisis de las sentencias SQL (número de JOIN) y de los planes de ejecución asociados (EXPLAIN ANALYZE: tiempo de ejecución y complejidad del árbol de ejecución).

Finalmente, para evaluar la capacidad de reflejar condicionantes y restricciones CoRIS, se planteó llevar a cabo la replicación de las cohortes de 3 estudios publicados basados en CoRIS/HICDEP v2018. Los trabajos fueron seleccionados por CoRIS en tres dominios combinados relevantes: presentación de eventos defintorios SIDA y mediciones de laboratorio (condition/measurement; Rava et al. PMID: [33882093](https://pubmed.ncbi.nlm.nih.gov/33882093/)); análisis de resultados en pautas de

antirretrovirales específicas (drug/measurement; Suárez et al. PMID: [32690099](#)); y análisis comparado entre pautas antirretrovirales (drug/measurement; Alejos et al. PMID: [32667674](#)).

5. Resultados

Se completó la implementación de la instancia CoRIS/OMOP, caracterizada según se muestra en la Tabla 1. Se identificaron 709 conceptos, de los cuáles 410 fueron conceptos relacionados específicamente con el ámbito de la salud (clínicos y epidemiológicos). Se muestra la distribución de conceptos: por vocabulario estándar (SNOMED-CT, RXNORM, ATC, UCUM, LOINC y otras); y por dominio: observación, condición de salud, medicación y tratamiento, medición, procedimiento; y otros. A continuación, se indica el número total de registros CoRIS/HICDEP, y en la nueva instancia CoRIS/OMOP. Finalmente, se indica el número total de registros por dominio en la instancia CoRIS/OMOP.

n° pac.	15.509					
concept.	709	410 *solo salud				
vocab.	snomed	rxnorm	atc	ucum	loinc	otro
	552	70	39	12	10	26
	77,86%	9,87%	5,50%	1,69%	1,41%	3,66%
*salud	253	70	39	12	10	26
	61,71%	17,07%	9,51%	2,93%	2,44%	6,34%
dominio	observ.	condit.	drug	meas.	proc.	otro
	355	158	116	43	7	30
	50,07%	22,28%	16,36%	6,06%	0,99%	4,30%
n° regist.	HICDEP		OMOP-CDM			
	2.981.379		11.646.605			
OMOP dominio	observ.	condit.	drug	meas.	proc.	otr.(5)
	8.251.784	66.919	70.252	3.231.531	3.990	22.129
	70,85%	0,57%	0,60%	27,75%	0,03%	0,19%

Tabla 1. Caracterización: vocabularios, conceptos, dominios

La Figura 1 muestra el resultado del examen de calidad de CoRIS/OMOP generada por Data Quality Assessment (DQD), que ofreció un resultado de prácticamente un 100% de tests superados satisfactoriamente (ratio: 2.776/2.781) en completitud, conformidad y plausibilidad. Los 5 errores detectados son consecuencia de inconsistencias acarreadas desde CoRIS/HICDEP; no será posible corregirlos porque no corresponde modificar la instancia origen.

	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	1911	0	1911	100%	285	2	287	99%	2196	2	2198	100%
Conformance	325	1	326	100%	54	1	55	98%	379	2	381	99%
Completeness	193	1	194	99%	8	0	8	100%	201	1	202	100%
Total	2429	2	2431	100%	347	3	350	99%	2776	5	2781	100%

Figura 1. Resultados calidad DQD para CoRIS/OMOP

En relación con la cobertura de conceptos, la Tabla 2 muestra los resultados referidos a los conceptos clínico-epidemiológicos CoRIS. Se muestra el grado de cobertura total y por dominio.

La evaluación de la cobertura de datos concluyó con una coincidencia del 100%, es decir, la instancia OMOP-CDM contiene todos los datos de origen (quedando corroborada simultáneamente la capacidad de extracción de datos).

	condit.	drug	meas.	proc.	person	event	Total
óptimo	143	74	33	11	31	32	324
	87,2%	59,2%	97%	78,6%	91,2%	82%	79%
sub-óptimo	21	49	1	2	3	7	83
	12,8%	39,2%	3%	14,3%	8,8%	18%	20,2%
no-mapeado	0	2	0	1	0	0	3
		1,6%		7,2%			0,8%
	164	125	34	14	34	39	410

Tabla 2. Conceptos mapeados (cobertura conceptos)

El resultado del análisis de las sentencias SQL se muestra en la Tabla 3, comparando para ambas instancias (promedio/ desviación estándar) el número de JOIN, y la complejidad de los planes de ejecución asociados.

	HICDEP	OMOP-CDM
n° JOIN	0,6 (±0,58)	1,08 (±1,08)
Plan de ejecución		
Nodos	4,24 (±1,69)	7,4 (±4,95)
Profundidad	3,64 (±1,15)	4,72 (±1,74)
Anchura	1,56 (±0,51)	1,76 (±0,52)
Tiempo ejecución (s)	0,28 (±0,61)	1,81 (±3,7)

Tabla 3. Análisis QUERY: complejidad, tiempos de ejecución

Para finalizar, la Tabla 4 muestra los resultados correspondientes a la replicación de las cohortes de los estudios seleccionados por CoRIS como indicador de capacidad para reflejar condicionantes y restricciones a través de CoRIS/OMOP y corroborando su capacidad para la replicación de evidencia. Para cada estudio se muestra el número de pacientes de la cohorte y sub-cohortes (n), reproducidas en cada instancia, y la diferencia en porcentaje. Para ello, se desarrollaron sentencias SQL y scripts R vinculados a cada estudio, respectivamente.

	HICDEP (n)	OMOP (n)	Dif (%)
Rava et al.	14.876	14.587	1,86
LP	6.636	6.225	1,93
LPAD	3.931	3.536	2,18
Suárez et al.	39	39	0
DRV+TDF	16	16	0
DRV+TAF	23	23	0
Alejos et al.	3.945	3.940	0,12
ABC/3TC/DTG	1.341	1.341	0
CBT/EVG/FTC/TAF	679	679	0
TDF/FTC+DTG	471	471	0
TDF/FTC/EVG/CBT	463	463	0
TDF/FTC/RPV	453	453	0
TDF/FTC+bDRV	329	324	1,5
TDF/FTC+RAL	209	209	0

Tabla 4. Resultado de replicación de cohortes/ sub-cohortes

6. Discusión

Los escenarios más habituales en el desarrollo de instancias para usos secundarios son: escenario de extracción que implica la conformación de una nueva instancia a partir de la localización-extracción de datos desde sistemas de información asistenciales; y, escenario de replicación, que plantea la generación de una nueva instancia por transformación de instancias previamente constituidas. La replicación conlleva la necesidad de demostrar objetivamente que se conservan las capacidades de la instancia origen (no-inferioridad). Éste último es el escenario abordado en este trabajo, en línea con otros como Voss [8] y Vaclav [9].

Es necesario señalar que CoRIS/HICDEP es una instancia para usos secundarios lo que, intrínsecamente, conlleva implicaciones como son: agregación/simplificación de conceptos; y alta densidad de conceptos en los registros.

En el primer caso, se generan situaciones de ambigüedad en origen que dificulta mapeo con terminologías estándar y por lo tanto la identificación precisa de conceptos estándar. En ambos casos, se acarrea pérdidas de información irreversibles ya que no existe posibilidad de acceso a las fuentes primarias. No obstante, HICDEP ha resultado suficientemente robusto y se ha alcanzado un mapeo semántico óptimo del 79% de los conceptos, resultado en línea con otras experiencias publicadas [8, 9]. El mapeo sub-óptimo del 20,2% (que implica cierto grado de ambigüedad), se centra esencialmente en el dominio DRUG y viene acarreado desde origen ya que en los tratamientos CoRIS no son registrados parámetros como dosis o vía de administración (simplificación). Los conceptos no mapeados son casi irrelevantes (0,8%) y se corresponde con fármacos no recogidos aún en terminologías estándar.

En segundo lugar, los modelos como HICDEP, plantean para los registros unos constructos extensos que aglutinan conceptos sin vinculación semántica, primando una orientación a explotación en ámbitos de interés concretos. Estas propuestas, minimizan el número de registros en los repositorios (alta densidad de conceptos por registro) y agilizan las consultas para extracción de datos, a costa de una menor flexibilidad (cambios menores implican modificaciones de los constructos). Por el contrario, el CDM OMOP plantea registros simples basados en conceptos únicos vinculados a su dominio natural, determinado por la semántica. Estas aproximaciones, generan más registros (menor densidad de conceptos por registro), haciendo más compleja la extracción de datos, pero son más flexibles para adoptar nuevas variables sin necesidad de cambios en el modelo. La multiplicación de registros se hace evidente en CoRIS/OMOP que cuadruplica el número de registros respecto a CoRIS/HICDEP.

La reproducción del informe CoRIS v2018 muestra que todos los datos están y son accesibles, demostrando la consistencia sintáctica y semántica que proporciona OMOP. Esta consistencia ha sido también corroborada en repositorios de otros ámbitos a través de la replicación de fenotipos consensuados [8, 9]. Finalmente, la fidelidad con la que han sido replicadas las cohortes experimentales de los trabajos analizados implica que es posible generar extracciones complejas de datos en OMOP-CDM y, por lo tanto, replicar evidencia. Tras análisis, las mínimas diferencias en los “n” de las cohortes fueron debidas a sutiles diferencias en el tratamiento aritmético de fechas y periodos temporales efectuados desde STATA, R o SQL.

Es evidente que el acceso SQL a CoRIS/OMOP es más costoso en complejidad de consultas y tiempo requerido que a CoRIS/HICDEP. En promedio, para el principal conjunto de consultas analizado (extracción masiva de datos para la replicación del informe CoRIS), el tiempo ha sido 6 veces superior en OMOP, aunque sobre un conjunto

de consultas no homogéneo, como se observa en las desviaciones estándar. No obstante, se puede afirmar que para consultas que implican menos de 2 operaciones JOIN, el tiempo de ejecución se reduce en OMOP pasa a ser 1,43 veces superior.

7. Conclusiones

OHDSI ofrece un amplio conjunto de herramientas abiertas y una activa comunidad profesional y científica que facilita la implementación y explotación sobre OMOP-CDM.

CoRIS/OMOP demuestra ser un repositorio no inferior a la CoRIS/HICDEP en los criterios establecidos (cobertura de conceptos; integridad de datos; replicación de cohortes complejas), poniendo de manifiesto que CoRIS/OMOP permite reproducir la evidencia generada desde CoRIS/HICDEP. Además, CoRIS/OMOP aporta las ventajas inherentes de los repositorios avanzados de datos en investigación biomédica y epidemiológica: interoperabilidad semántica, y escalabilidad en capacidad de integración de nuevos ámbitos y fuentes de datos.

Agradecimientos

Trabajos realizados en el marco del proyecto AES, Smart-PITeS (PI18CIII/ 00004, PI18/ 01047); y proyecto RIS EPICLIN15_2019, como contribución a RETIC RIS RD16CIII/0002/0006.

Referencias

- [1] eHealth: Digital health and care. European health-data space. https://ec.europa.eu/health/ehealth/dataspace_en (Consultada: junio 2022)
- [2] Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc.* 2013;20(1):144-151. doi:10.1136/amiainl-2011-000681
- [3] Kent, S., Burn, E., Dawoud, D. et al. Common Problems, Common Data Model Solutions: Evidence Generation for Health Technology Assessment. *PharmacoEconomics*, vol 39, 2021, pp 275–285 (DOI: 10.1007/s40273-020-00981-9)
- [4] Observational Medical Outcomes Partnership (OMOP) CDM. <https://www.ohdsi.org/data-standardization/the-common-data-model/> (Consultada: junio 2022)
- [5] CoRIS: Cohorte de la Red Española de Investigación en SIDA. <https://coris.isciii.es/> (Consultada: junio 2022)
- [6] HICDEP (HIV Cohorts Data Exchange Protocol). <https://hicdep.org> (Consultada: junio 2022)
- [7] Garza Met al. Evaluating common data models for use with a longitudinal community registry. *J Biomed Inform.* 2016 Dec; 64: 333-341. doi: 10.1016/j.jbi.2016.10.016. Epub 2016 Oct 29. PMID: 27989817; PMCID: PMC6810649
- [8] Voss EA et al. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. *J Am Med Inform Assoc.* 2015 May;22(3):553-64. doi: 10.1093/jamia/ocu023. Epub 2015 Feb 10. PMID: 25670757; PMCID: PMC4457111.
- [9] Vaclav Papez, et al, Transforming and evaluating electronic health record disease phenotyping algorithms using the OMOP common data model: a case study in heart failure, *JAMIA Open*, Volume 4, Issue 3, July 2021, ooab001, <https://doi.org/10.1093/jamiaopen/ooab001>