

XXIV Congreso Nacional de Informática de la Salud

2021

Infors@Iud

"La Evolución del ecosistema digital en Salud"

Organiza



Hotel NH Collection Madrid Eurobuilding
c/ Alberto Alcocer 8
28036 Madrid

MADRID - 21, 22 y 23 de junio

LIBRO DE COMUNICACIONES, PÓSTERS Y
PROYECTOS DE INNOVACIÓN



Síguenos en twitter:
@SEISeSalud

SOCIO TECNOLÓGICO PRINCIPAL

SOCIO TECNOLÓGICO COLABORADOR



Primera Edición, Junio 2021

Queda rigurosamente prohibida, sin la autorización escrita de los titulares del “copyright”, bajo las sanciones establecidas por las leyes, la reproducción parcial o total de esta obra por cualquier medio o procedimiento, comprendidos la reprografía y el tratamiento informático y la distribución de ejemplares de ella mediante alquiler o préstamos públicos.

SEIS, Sociedad Española de Informática de la Salud

<http://www.seis.es>

Secretaría Técnica, CEFIC S.L
C/ Enrique Larreta nº 5 Bajo izquierda
28036 MADRID Telf. : 913889478

Printed in Spain-Impreso en España

ISBN:978-84-09-31501-7

Editor General:
Sáez Ayerra, Luciano

Editor Científico:
Monteagudo Peña, José Luis

Editor técnico:
Alonso Suero, Elvira

COMITES

COMITÉ ORGANIZADOR

Presidente
Luciano Sáez Ayerra

Coordinadora General
Elvira Alonso Suero

COMITÉ DE GESTIÓN

Martín Begoña Oleaga
Francisco Martínez del Cerro
José Quintela Seoane

COMITÉ DE PROGRAMA

Coordinador
Juan Fernando Muñoz Montalvo

Miembros
Carlos García Codina
Francisco Martínez del Cerro
Guillermo Vázquez González

COMITÉ CIENTÍFICO

Presidente
José Luis Monteagudo Peña

Área Institucional

Coordinadores
Carlos García Codina
Martín Begoña Oleaga

Miembros
E. Bausa Llamazares
M. Begoña Oleaga
M. A. Cisneros Martín
J.I. Coll Clavero
M.L. De los Mártires
M. Escudero Sánchez
S. García Blanco
G. M. Garmendia Pérez
J. A. Gómez Palomeque
E. Linares Morcillo
Y. Lopez Mínguez
A. I. Martínez Albarrán
M. J. Millán Muñoz
C. Moreno Gutiérrez
J. F. Muñoz Montalvo
P. Pérez i Sust
B. Rosón Calvo
M. Salvatierra Baladrón

Área Profesional

Coordinadores
Guillermo Vázquez González
Javier López Caverro

Miembros
L. M. Béjar Prado
A. Blanco Rubio
S. Carreres Montell
M. Chavarría Díaz
M. Estupiñán Ramírez
C. Gallego Pérez
A. Gómez Lafón
C. M. Hernández Marín
C. Hernández Salvador
A. Martínez Aparisi
L. Morell Baladrón
A. Muñoz Carrero
C. Parra Calderón
A. Peña González
A. Poncel Falcó
J. Reig Redondo
J. Sacristán París
F. Sánchez Laguna

Área Enfermería

Coordinadora
I. Moro Casuso
Miembros
C. González del Pino
C. Moreno -Chocano Gutiérrez
N. Moro Tejedor

Área Farmacia

Coordinadora
L. Fidalgo García
Miembros
C. Carmona
J. L. Ceruelo Bermejo
R. Fonseca Álvarez
M^a. A. Giménez Ferrer
A. T. López Navarro
L. Muñoz Fernández
F. Pérez Hernández
A. Smits Cuberes

Área Tecnológica

Coordinadores
Francisco Martínez del Cerro
José Sacristán París

Miembros
D. Álvarez López
M. Felipe Gil
B. Lobato Beiro
I. López Parrilla
M. A. Montero
J. María Moyano
C. Piqueras Picón
O. Sanz Martín

OBTENCIÓN DE UN CONJUNTO DE DATOS DERIVADOS DE LA HCE PARA INVESTIGACIÓN EN VIH: PRUEBA DE CONCEPTO

P. RUBIO MAYO¹, M. PEDRERA JIMÉNEZ², N. GARCÍA BARRIO², O. BISBAL PARDO², J. CRUZ ROJO², P. SERRANO BALAZOTE², M. PASCUAL CARRASCO³, J.L. BERNAL SOBRINO²

¹Instituto de Investigación Sanitaria del Hospital 12 de Octubre, Madrid, España.

²Hospital Universitario 12 de Octubre, Madrid, España.

³Instituto de Investigación Carlos III (ISCIII), Madrid, España.

INTRODUCCIÓN

La Historia Clínica Electrónica (HCE) puede definirse como un repositorio de datos sanitarios en formato digital que son almacenados e intercambiados de manera segura, y accesibles por múltiples usuarios autorizados, siendo fundamental en la práctica asistencial y en otros usos denominados como secundarios, donde se incluye la investigación sanitaria [1,2]. Para aprovechar al máximo la capacidad de la HCE, es necesario que los datos puedan ser intercambiados y combinados sin pérdida de significado, conocido como interoperabilidad semántica [3]. En investigación sanitaria, esto permitiría reutilizar los datos asistenciales, mejorando el escenario actual donde cada propósito de análisis requiere de un sistema específico paralelo a la HCE [4].

Esto se aplicaría al Virus de la Inmunodeficiencia Humana (VIH), uno de los problemas de salud más graves a nivel mundial. Al cierre del 2019, 38 millones de personas vivían con la infección del VIH y hubo 1,7 millones de nuevos contagios. En España, se estima que hay entre 120.000 y 150.000 personas vivas que padecen la infección de VIH y más de una cuarta parte permanece sin diagnosticar [5]. A 31 de marzo de 2021 se tiene constancia de que aproximadamente 21.600 personas residentes en la Comunidad de Madrid presentaban infección por VIH [6]. Hoy en día, gracias a los avances médicos y al tratamiento con antirretrovirales, las personas con VIH pueden vivir sin sufrir los síntomas. No obstante, el virus VIH y el sida siguen siendo un problema de salud pública, y la gestión e investigación sobre esta enfermedad requiere disponer de datos de salud de calidad [4].

Por ello, en 2003 se constituyó en España la Red Española de Investigación en Sida (RIS), una estructura coordinada y organizada para la investigación en VIH/sida. En este marco, se creó una cohorte prospectiva multicéntrica de pacientes VIH-positivos, conocida como CoRIS, de la que se recogen datos de manera coordinada, tanto epidemiológica como demográfica [7]. Actualmente participan 42 centros, entre los que se encuentra el Hospital Universitario 12 de Octubre. Estos datos relativos a pacientes de VIH para CoRIS se registran de forma manual en un cuaderno de recogida (CRD), suponiendo un esfuerzo de registro datos en un sistema paralelo a los asistenciales.

OBJETIVO

En este trabajo se pretende estandarizar y automatizar el proceso de obtención de los datos requeridos por el modelo de CoRIS, aprovechando los datos ya registrados en la HCE durante la práctica asistencial, transformándolos de acuerdo con las restricciones de dicho modelo de uso secundario.

MATERIAL Y MÉTODOS

Para este trabajo se ha aplicado una metodología desarrollada en trabajos previos [4], aplicada a una prueba de concepto de extracción y transformación de datos desde la HCE al modelo CoRIS. Para ello, en primer lugar, se ha estudiado el set de datos requerido por CoRIS y se han seleccionado un subconjunto de sus elementos. En segundo lugar, estos elementos se han modelado en la HCE del Hospital, estandarizándolos de acuerdo a

especificaciones internacionales como ISO-13606 [8], SNOMED CT [9] y LOINC [10]. Por último, se han definido e implementado las reglas de transformación necesarias para generar el modelo CoRIS.

1.1. Identificación del conjunto de datos

CoRIS es un modelo de datos donde se recoge información de los pacientes de VIH a lo largo del tiempo [7]. Su objetivo es crear un repositorio colaborativo entre investigadores que permita el análisis de la información epidemiológica, de evolución de los pacientes, así como de respuesta a tratamientos antirretrovirales y otros factores, con el fin de aumentar el conocimiento sobre esta patología. Se ha analizado este modelo y se ha seleccionado un subconjunto de variables registradas en CoRIS. Dichas variables se corresponden con 9 observaciones clínicas generales (talla, peso, presión arterial y hábito tabáquico) y 14 observaciones de laboratorio, como las pruebas diagnósticas de VIH o el hemograma.

1.2. Modelado y estandarización de la información

El modelado y la formalización de la información se ha realizado de acuerdo a la norma ISO-13606 [8]. Por un lado, la norma ISO-13606, basada en el paradigma dual de modelo de referencia y modelo de arquetipos, se ha empleado para definir el modelo de conceptos que permita la recogida homogénea y controlada de los datos y su posterior uso. Se ha empleado este estándar al ser adoptado por el Ministerio de Sanidad para la construcción de documentos clínicos interoperables y estar implantado en el Hospital Universitario 12 de Octubre [4].

Para modelar el conjunto de variables seleccionadas de acuerdo al modelo de referencia, ha sido necesario utilizar los siguientes tipos de datos:

- Cantidad Física (PQ): para variables que reportan un valor numérico con unidad de medida
- Valor codificado (CV): para variables que reportan un conjunto de posibles datos codificados
- Entero (Integer): para variables que reportan un valor entero
- Fecha (Date): para variables que reportan una fecha

El enlace semántico se ha realizado con SNOMED CT para las entidades observables de carácter clínico [9], y LOINC para las entidades observables obtenidas de laboratorio [10]. La tabla 1 muestra el conjunto de conceptos codificados de acuerdo con los estándares previamente explicados.

Tabla 1. Conjunto de entidades observables estandarizadas para VIH

Elemento	Tipo de dato	Unidades / Diccionario	Enlace semántico
Peso	PQ	Kg	27113001 peso corporal (entidad observable)
Talla	PQ	cm	50373000 altura corporal (entidad observable)
Presión sistólica	PQ	mmHg	271649006 presión sistólica (entidad observable)
Presión diastólica	PQ	mmHg	271650006 presión diastólica (entidad observable)
Año de inicio del hábito tabáquico	Date		266929003 tabaquismo iniciado (hallazgo)
Año abandono del hábito tabáquico	Date		160625004 fecha en que dejó de fumar (entidad observable)
Índice tabáquico acumulado	PQ	Paquetes/año	782516008 cantidad de paquetes de cigarrillos fumados por día por el número de años en que la persona ha fumado para medir la exposición al tabaco durante un período de tiempo (entidad observable)
Cantidad cigarros al día	PQ	Cigarros/día	230057008 consumo de cigarros (entidad observable)
Hábito tabáquico	CV	Sí No	266918002 tabaquismo (entidad observable)
Valor absoluto de los CD4	PQ	células/mm3	24467-3 procedimiento de recuento absoluto de células CD4 (procedimiento)
Porcentaje de CD4	PQ	%	8128-1 Linfocitos CD4 /100 células en sangre
Valor absoluto de CD8	PQ	células/mm3	14135-8 recuento absoluto de linfocitos T CD8 (procedimiento)
Porcentaje de CD8	PQ	%	80222-3 Linfocitos CD8/100 células en sangre
Carga viral	PQ	copias/mL	315124004 prueba de carga viral de virus de la inmunodeficiencia humana (procedimiento)
Hemoglobina	PQ	g/dL	718-7 Hemoglobina (masa/volumen) en sangre
Plaquetas	PQ	g/dL	777-3 Plaquetas (#/masa) en sangre por conteo automático
Creatinina	PQ	mg/dL	2160-0 Creatinina (masa/volumen) en suero o volumen
Albumina	PQ	g/dL	1751-7 Albumina (masa/volumen) en suero o plasma
AST	PQ	UI/L	1920-8 Aspartato aminotransferasa (actividad enzimática/volumen) en suero o plasma
ALT	PQ	UI/L	1742-6 Alanina aminotransferasa (actividad enzimática/volumen) en suero o plasma
Colesterol Total	PQ	mg/dL	2093-3 Colesterol
Colesterol HDL	PQ	mg/dL	2085-9 HDL-Colesterol
Triglicéridos	PQ	mg/dL	2571-8 Triglicéridos (masa/volumen) en suero o plasma

1.3. Definición de reglas de transformación

Una vez definido el catálogo de conceptos estándares, se han definido las reglas de equivalencia entre estos conceptos y los solicitados en el modelo CoRIS. Con esto, se ha implementado un algoritmo para generar dicho modelo de uso secundario, utilizando el lenguaje de programación R para procesar y transformar los datos extraídos desde la HCE [11]. Respecto al algoritmo de transformación, este parte de una extracción bruta de datos desde la HCE de los elementos definidos y para la cohorte seleccionada. A continuación, sobre esta información se aplican las diferentes reglas definidas en el apartado anterior y se obtienen las variables requeridas en el modelo de explotación CoRIS.

La figura 1 muestra un ejemplo de transformación sobre el concepto de CoRIS “hábito tabáquico” de valores “fumador”, “exfumador” y “no fumador”, inferido a partir de una variable de la HCE ‘Fumador’ (con respuestas ‘Sí’ y ‘No’), ‘Paquetes al año’ y ‘Fecha fin de hábito tabáquico. Así, si existían registros anteriores de fumador para aquellos en los que la variable tomaba valor ‘No’ actualmente, se infería ‘Ex fumador’. También se ha tenido en cuenta si el número de cigarrillos diarios era menor a 1 al día para inferir en ‘fumador ocasional’, partiendo de los paquetes al año registrados ((Paquetes al año×20)/365). Por último, si los pacientes tenían la variable ‘Fecha fin de hábito’ rellena, se infería como ‘Ex fumador’.

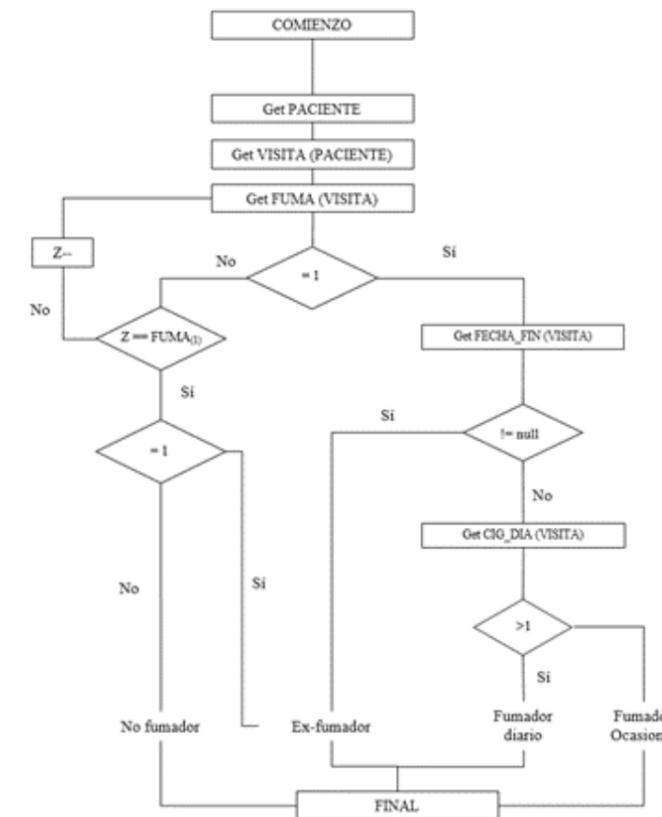


Figura 1: Algoritmo de transformación y extracción de datos relacionados con el hábito tabáquico

RESULTADOS

Los resultados obtenidos hasta la fecha se corresponden con el catálogo estándar de las entidades observables escogidas para la prueba de concepto descrita. Así mismo, se ha diseñado un algoritmo de transformación para inferir las variables del modelo de uso secundario en base a estos conceptos seleccionados y normalizados. Por último, se ha aplicado el algoritmo a un subconjunto de la cohorte VIH de 722 pacientes.

La Tabla 3 muestra el volumen de datos que se obtendría actualmente para cada una de las variables seleccionadas. Concretamente, en la primera columna se muestra el número de registros existentes en la HCE para cada elemento, mientras que, en la segunda columna se refleja el número de registros resultante tras realizar las transformaciones pertinentes para generar el modelo de uso secundario de CoRIS. Adicionalmente, se incluyen dos columnas para indicar el número de pacientes que tienen recogidos datos para cada concepto y el porcentaje de los mismos respecto al total de miembros de la cohorte (722 pacientes).

Tabla 2. Dataset de CoRIS generada a partir de la HistoriaClínico Electrónica (HCE)

Concepto	HCE (N)	CoRIS (N)	Pacientes (N)	Pacientes (%)
Altura	9053	9053	707	98%
Peso	9492	9476	710	98%
Fumador	308	308	117	16%
Cigarros/Día	71	70	55	8%
Cigarros/Semana	71	70	55	8%
Edad/Año inicio	35	33	23	3%
Edad/Año fin	47	42	31	4%
P. Sistólica	20150	20150	704	98%
P. Diastólica	20138	20138	704	98%
NCD4	9111	9086	701	97%
%CD4	9079	9076	701	97%
NCD8	9113	9113	701	97%
%CD8	9085	9085	701	97%
Carga viral	2541	2516	704	98%
Carga Viral Bajo limite	2541	2516	704	98%
Carga Viral Sobre limite	2541	2516	704	98%
Hemoglobina	18311	18289	708	98%
Plaquetas	18255	18245	708	98%
Creatinina	17205	17173	710	98%
Albumina	15570	15541	707	98%
AST	16247	16203	706	98%
ALT	16393	16361	707	98%
Colesterol	11263	11248	699	97%
HDL	10278	10268	698	97%
Triglicéridos	11329	11315	699	97%

Se puede observar que variables recogidas en la HCE de manera habitual, como pruebas de laboratorio básicas o la antropometría, tienen una cobertura cercana al 100% de la cohorte utilizada en la prueba de concepto. Sin embargo, variables que no se recogen, o se hace comunmente en texto libre, como las relativas a los hábitos de vida, tienen una cobertura baja, en torno al 10%. Esto es una característica de los proyectos con datos de HCE, donde solo es posible obtener aquellos datos que el profesional asistencial registró, de manera estructurada, en la asistencia del paciente, y remarca la necesidad de mejorar los procesos de registro en los sistemas asistenciales

DISCUSIÓN

El modelo de referencia de la norma ISO-13606 ha permitido modelar los conceptos clínicos y, a su vez, el modelo de arquetipo ha permitido su formalización y la definición del enlace semántico mediante las terminologías SNOMED CT y LOINC. El uso de estándares ha permitido crear un proceso flexible y escalable a otras organizaciones y condiciones de salud [4].

AGRADECIMIENTOS

Este estudio ha sido subvencionado por el Instituto de Salud Carlos III mediante los proyectos "PI18/01047" y "PI18CIII0004" (Cofinanciado por el Fondo Europeo de Desarrollo Regional/Fondo Social Europeo. "Una manera de hacer Europa"/ "El FSE invierte en tu futuro").

BIBLIOGRAFÍA

- [1] Häyrynen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: A review of the research literature. *Int J Med Inform* 2008;77:291–304. doi:10.1016/j.ijmedinf.2007.09.001.
- [2] Safran C, Bloomrosen M, Hammond E, et al. Toward a National Framework for the Secondary Use of Health. *J Am Med Informatics Assoc* 2007;14:1–9. doi:10.1197/jamia.M2273.
- [3] Sun H, Depraetere K, De Roo J, et al. Semantic processing of EHR data for clinical research. *J Biomed Inform* 2015;58:247–59. doi:10.1016/j.jbi.2015.10.009.
- [4] Pedrera-Jiménez M, García-Barrio N, Cruz-Rojo J, et al. Obtaining EHR-derived datasets for COVID-19 research within a short time: a flexible methodology based on Detailed Clinical Models. *J Biomed Inform*. 2021;115:103697. doi:10.1016/j.jbi.2021.103697.
- [5] ISCIII, "Vigilancia epidemiológica del VIH y Sida en España 2019. Sistema de información sobre nuevos diagnósticos de VIH. Registro Nacional de casos de Sida - D.G de Salud Pública / Centro Nacional de Epidemiología" 2020.
- [6] R. D. V. Epidemiológica, "INFORME EPIDEMIOLÓGICO VIGILANCIA VIH / SIDA Febrero 2021," vol. 2021, pp. 2–6, 2021.
- [7] Caro-Murillo AM, Castilla J, Pérez-Hoyos S, et al. Cohorte RIS de pacientes con infección por VIH sin tratamiento antirretroviral previo (CoRIS): metodología y primeros resultados. *Enferm Infecc Microbiol Clin*. 2007;25(1):23-31. doi:10.1157/13096749.
- [8] Santos MR, Bax MP, Kalra D. Building a logical EHR architecture based on ISO 13606 standard and semantic web technologies. *Stud Health Technol Inform* 2010;160:161–5. doi:10.3233/978-1-60750-588-4-161.
- [9] Donnelly K. SNOMED-CT: The advanced terminology and coding system for eHealth. *Stud Health Technol Inform*. 2006;121:279-290.
- [10] McDonald CJ, Huff SM, Suico JG, et al. LOINC, a universal standard for identifying laboratory observations: A 5-year update. *Clin Chem* 2003;49:624–33. doi:10.1373/49.4.624.
- [11] R Foundation for Statistical Computing. <https://www.r-project.org/about.html>. Accedido: 11 de junio de 2021.