



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Programa de doctorado de tecnologías para la salud y el bienestar

Enriquecimiento de la historia clínica electrónica con información de sistemas de ayuda a la decisión clínica y datos enlazados abiertos

Tesis Doctoral

Autor

Alejandro Mañas García

Directores

José Alberto Maldonado Segura

María del Mar Marcos López

Tutora

Montserrat Robles Viejo

Valencia, julio 2022

Agradecimientos

Esta tesis ha sido posible gracias al apoyo incondicional de mi familia. Mis padres, Marcelino y María Victoria, y mi hermana Elena, cuyos ánimos han sido fundamentales en los momentos más difíciles. Gracias también a mi pareja, mejor amiga y compañera de viaje, Tatiana, por haber recorrido a mi lado un largo camino de ocho años de investigación, desarrollo y transferencia tecnológica, que quedan plasmados en este documento. Estoy enormemente agradecido con mis directores y tutores de tesis, Mar, Alberto y Montse, cuya profesionalidad, experiencia académica y rigor científico, han permitido alcanzar unos resultados de investigación más que satisfactorios, reconocidos en prestigiosas revistas científicas del área. Ellos han sido la luz que ha iluminado el camino en busca de la excelencia académica. Gracias por todo lo aprendido, por vuestro conocimiento y vuestra exigencia. Por último, pero no menos importante, agradezco sinceramente toda la experiencia adquirida y compartida con los compañeros de Quibim. Con ellos, hemos dado un paso más allá en esta tesis. Además del trabajo de investigación y las publicaciones científicas, esta tesis presenta un ejercicio de transferencia tecnológica. Ha sido una experiencia muy enriquecedora ver los frutos de esta investigación funcionando en hospitales y contribuyendo al proceso asistencial. A los mencionados anteriormente, y a tantos otros que dejo por mencionar, gracias por todo. Esta tesis también es vuestra.

Contenido

Agradecimientos	2
Contenido	4
Lista de figuras	8
Lista de tablas	11
Abreviaturas y acrónimos	13
Glosario	17
Resumen	25
Resum	26
Summary	27
Capítulo 1. Introducción	29
1.1. Motivación	29
1.2. Hipótesis, preguntas de investigación y objetivos	31
1.2.1. Hipótesis	31
1.2.2. Preguntas de investigación y objetivos	31
1.3. Contribuciones	33
1.4. Estructura de la tesis	40
Capítulo 2. Estado del arte	44
2.1. Sistemas de información sanitarios	45
2.2. Historia clínica electrónica	48
2.3. Estándares de historia clínica electrónica	48
2.3.1. ISO13606	50
2.3.2. OpenEHR	52
2.3.3. HL7 CDA	53
2.3.4. HL7 FHIR	53
2.4. Estándares de imagen médica	53
2.5. SADC basados en guías clínicas y biomarcadores de imagen	56
2.5.1. SADC basados en reglas	59
2.5.2. SADC basados en visión por computación	59
2.5.3. SADC basados en aprendizaje automático	61
2.6. Informe estructurado radiológico	61
2.7. Interoperabilidad de sistemas de información sanitarios	66
2.8. Soluciones interoperables de integración de sistemas de HCE con resultados de SADC y datos enlazados abiertos	69
Capítulo 3. Historia Clínica Electrónica aumentada con datos enlazados abiertos	75

3.1. Introducción	75
3.2. Planteamiento	76
3.3. Metodología	82
3.3.1. Identificación del contenido de aumento para un arquetipo	83
3.3.2. Transformación de las EDI en una representación semántica	84
3.3.3. Consulta de conjuntos de datos de la web semántica	87
3.3.4. Transformación del contenido de aumento en un extracto de HCE normalizado	90
3.3.5. Integración del contenido de aumento normalizado y el extracto inicial de la HCE	92
3.4. Resultados	95
3.5. Caso de uso	98
3.6. Discusión	107
Capítulo 4. Informe radiológico estructurado enriquecido con sistemas de ayuda a la decisión clínica	110
4.1. Introducción	110
4.2. Metodología	112
4.3. Resultados	116
4.4. Discusión	122
Capítulo 5. Discusión	126
Capítulo 6. Conclusiones y trabajo futuro	134
6.1. Conclusiones	134
6.2. Trabajo futuro	137
Bibliografía	140
Anexo I. Plantillas de informe radiológico estructurado	154
Resonancia magnética de cerebro	155
Resonancia magnética de columna completa sin contraste	156
Resonancia magnética de rodilla	157
Resonancia magnética de cadera	158
Tomografía computarizada de abdomen y pelvis	159
Tomografía computarizada de cuello	160
Tomografía computarizada de huesos faciales	161
Tomografía computarizada de senos paranasales	162
Tomografía computarizada de embolia pulmonar de tórax	163
Radiografía de rodilla	164
Radiografía de tórax con dos proyecciones	165

Lista de figuras

Figura 1. Estructura de la tesis. Orden y relaciones entre capítulos y anexos.	41
Figura 2. Diagrama de componentes de los sistemas de información sanitarios de HCE y radiología.	46
Figura 3. Diagrama de secuencia de los flujos de trabajo de citación para prueba de imagen médica, informado radiológico y visualización de HCE e informes radiológicos.	47
Figura 4. Componentes más relevantes del MR de la norma ISO13606.....	52
Figura 5. Jerarquía de niveles de información de los objetos DICOM.	55
Figura 6. Modelo entidad-relación de DICOM SR.	64
Figura 7. Ejemplo de plantilla de IRE de acuerdo al modelo de referencia DICOM-SR.....	65
Figura 8. Operaciones de intercambio de datos para construir extractos de HCE aumentados.	79
Figura 9. Metodología para definir el arquetipo de aumento (pasos 2 a 6) y el proceso completo para obtener el arquetipo aumentado (pasos 1 a 7).....	82
Figura 10. La consulta SPARQL y el grafo de consulta para obtener el contenido de aumento.	88
Figura 11. La consulta SPARQL y el grafo de consulta del ejemplo de la sección 3.3.3.	89
Figura 12. Resultados de la consulta SPARQL del ejemplo de la sección 3.3.3 formateados en XML, según el esquema canónico del W3C para los resultados de las consultas.	89
Figura 13. Visualización en LinkEHR del arquetipo de aumento CEN-EN13606-CLUSTER.DrugDrugInteraction.v1 y representación del código HTML de una instancia del arquetipo de aumento CEN-EN13606-CLUSTER.DrugDrugInteraction.v1.	92
Figura 14. Visualización de arquetipos en LinkEHR. A la izquierda, el arquetipo inicial CEN-EN13606-ENTRY.Medicacion.v1. A la derecha, el arquetipo aumentado CEN-EN13606-ENTRY.MedicacionAumentada.v1.....	95
Figura 15. Descripción del proceso para obtener un extracto de HCE aumentada.....	96
Figura 16. Visualización mediante LinkEHR del arquetipo inicial CEN-EN13606-COMPOSITION.HistoriaClinicaResumida.v1 y el extracto inicial de HCE.....	99
Figura 17. Consultas SPARQL para obtener los contenidos de aumento..	103
Figura 18. Representación de arquetipos de aumento en LinkEHR..	104
Figura 19. Representación de arquetipo inicial y arquetipo aumentado en LinkEHR.....	104
Figura 20. Representación HTML del contenido del aumento..	106
Figura 21. Representación HTML del extracto de HCE resumida aumentada.....	106
Figura 22. Flujo de trabajo del informado de un caso con el sistema de IRE enriquecido..	115
Figura 23. Ejemplo de cómo la información relevante del informe procedente de distintas fuentes se combina en un extracto de texto en lenguaje natural.	116

Figura 24. Una sección de la plantilla de IRE de IRM de próstata relacionada con el tamaño y el volumen de la próstata.....	117
Figura 25. Validadores de campos y reglas lógicas que cambian dinámicamente el formulario del IRE..	118
Figura 26. Ciclo de vida de la plantilla de IRE conforme al perfil de integración IHE MRRT..	119
Figura 27. IRE de IRM de próstata, sección PI-RADS..	121
Figura 28. IRE de IRM de próstata, integrado con una plataforma de biomarcadores de imagen..	121
Figura 29. Arquetipo de informe de resultados de pruebas de imagen del CMDIC.	131
Figura 30. Formulario de IRE para IRM de cerebro.....	155
Figura 31. Formulario de IRE para IRM de columna completa sin contraste..	156
Figura 32. Formulario de IRE para IRM de rodilla..	157
Figura 33. Formulario de IRE para IRM de cadera.....	158
Figura 34. Formulario de IRE para TAC de abdomen y pelvis.....	159
Figura 35. Formulario de IRE para TAC de cuello.....	160
Figura 36. Formulario de IRE para TAC de huesos faciales..	161
Figura 37. Formulario de IRE para TAC de senos paranasales.....	162
Figura 38. Formulario de IRE para TAC de embolia pulmonar de tórax.....	163
Figura 39. Formulario de IRE para radiografía de rodilla..	164
Figura 40. Formulario de IRE para radiografía de tórax con dos proyecciones.....	165

Lista de tablas

Tabla 1. Estructura de la tesis y relación de capítulos con preguntas de investigación, objetivos y publicaciones.	42
Tabla 2. Identificador, nombre y descripción de etiquetas públicas de la cabecera DICOM ampliamente utilizadas.	55
Tabla 3. Resumen de nueve sistemas de datos e informado RADS (Reporting And Data System) publicados por el ACR.....	58
Tabla 4. Resultados de la consulta SPARQL del ejemplo de la sección 3.3.3.....	90
Tabla 5. Entradas y salidas del procedimiento automático para construir las tripletas de enlace EDI para el aumento de HCE de las interacciones medicamentosas.....	100
Tabla 6. Entradas y salidas del procedimiento automático para construir las tripletas de enlace EDI para el aumento de HCE de tratamientos relacionados con episodios activos del paciente.	101
Tabla 7. Tripletas de enlace EDI generadas para cada aumento de HCE, a partir del extracto inicial de HCE introducido en la figura 16 y los argumentos descritos en la tabla 5 y la tabla 6, respectivamente.....	102
Tabla 8. Esquemas de datos involucrados en los mapeos para generar instancias de la HCE resumida aumentada.	105

Abreviaturas y acrónimos

- ADC:** Coeficiente de Difusión Aparente. Del inglés: “Apparent Diffusion Coefficient”.
- ADL:** Lenguaje de definición de arquetipos. Del inglés “Archetype Definition Language”.
- BI:** Inteligencia de negocio. Del inglés “Business Intelligence”.
- BI-RADS:** Sistema de datos e informado de imágenes mamarias. Del inglés: “Breast Imaging-Reporting and Data System”.
- C-RADS:** Sistema de datos e informado de imágenes de colonoscopia virtual. Del inglés “Colonography Reporting and Data System”.
- CA:** Contenido de Aumento.
- CAD-RADS:** Sistema de datos e informado de imágenes de enfermedades arteriales coronarias. Del inglés “Coronary Artery Disease Reporting and Data System”.
- CEM:** Modelo de elementos clínicos. Del inglés “Clinical Elements Model”.
- CMDIC:** Conjunto Mínimo De Informes Clínicos.
- CNN:** Redes neuronales convolucionales. Del inglés “Convolutional Neural Networks”.
- DCE:** Secuencia de IRM de perfusión con contraste mejorado. Del inglés “Dynamic Contrast Enhanced”.
- DICOM:** Estándar de imagen médica y comunicación. Del inglés “Digital Imaging and COunication in Medicine”.
- DICOM-SR:** Estándar de informe estructurado DICOM. Del inglés “DICOM - Structured Report”.
- DL:** Lesión dominante. Del inglés “Dominant lesion”.
- DWI:** Secuencia de IRM de difusión ponderada. Del inglés “Weighted Diffusion Image”.
- ED:** Esquema Destino.
- EDE:** Entidades de datos equivalentes.
- EDI:** Entidades de interés.
- EF:** Esquema Fuente.
- GCI:** Guías Clínicas Informatizadas.
- HCE:** Historia Clínica Electrónica.
- HCDSNS:** Historia Clínica Digital del Sistema Nacional de Salud.
- HIS:** Sistema de información sanitaria. Del inglés: “Health Information System”.
- HIMSS:** Sociedad de sistemas de información y gestión sanitaria norteamericana. Del inglés “Healthcare Information and Management Systems Society”.
- HL7:** Organización americana para la normalización de información cínica. Del inglés: “Health Level 7”.
- HTML:** Lenguaje de maquetado de hipertexto. Del inglés: “Hyper-Text Markup Language”.
- IA:** Inteligencia Artificial.

IEEE: Instituto de Ingenieros Eléctricos y Electrónicos. Del inglés “Institute of Electrical and Electronics Engineers”.

IHE: Integración de la empresa sanitaria. Del inglés “Integrating the Healthcare Enterprise”.

IRE: Informe Radiológico Estructurado.

IRM: Imagen de Resonancia Magnética.

JSON: Notación de objetos de JavaScript. Del inglés: “JavaScript Object Notation”.

LD: Datos enlazados. Del inglés “Linked Data”.

LI-RADS: Sistema de datos e informado de imágenes hepáticas. Del inglés “Liver Imaging Reporting and Data System”.

LOD: Datos enlazados abiertos. Del inglés: “Linked Open Data”.

LODC: Nube de datos enlazados abiertos. Del inglés: “Linked Open Data Cloud”.

Lung-RADS: Sistema de datos e informado de imágenes de pulmón. Del inglés “Lung Reporting and Data System”.

MA: Modelo de Arquetipos.

MEAN: Stack tecnológico que hace referencia a las tecnologías: MongoDB, ExpressJS, Angular, NodeJS.

MR: Modelo de Referencia.

MRRT: Gestión de plantillas de informe radiológico. Del inglés: “Management of Radiology Reporting Templates”.

NDFRT: Terminología de referencia del archivo nacional de fármacos. Del inglés “National Drug File - Reference Terminology”.

NI-RADS: Sistema de datos e informado de imágenes de cuello. Del inglés “Neck Imaging Reporting And Data System”.

O-RADS: Sistema de datos e informado de imágenes de ovario. Del inglés “Ovarian-Adnexal Imaging Reporting Data System”.

ODE: Organizaciones de Desarrollo de Estándares.

OWL: Lenguaje de ontologías web. Del inglés “Ontology Web Language”.

PACS: Sistema de archivo y comunicación de imágenes. Del inglés “Picture Archiving and Communication System”.

PET: Tomografía por emisión de positrones. Del inglés “Positron Emission Tomography”.

PI-RADS: Sistema de datos e informado de imágenes de próstata. Del inglés “Prostate Imaging Reporting And Data System”.

RDF: Marco de descripción de recursos. Del inglés “Resource Description Framework”.

RDFS: Esquema de RDF. Del inglés “Resource Description Framework Schema”.

RECIST: Criterios de evaluación de respuesta en tumores sólidos. Del inglés “Response evaluation criteria in solid tumors”.

RIS: Sistema de información radiológica. Del inglés “Radiological Information System”.

RSNA: Sociedad de radiología de Norteamérica. Del inglés “Radiological Society of North America”.

SADC: Sistemas de Ayuda a la Decisión Clínica.

SCP: Proveedor de clases de almacenamiento. Del inglés “Storage Class Provider”.

SIS: Sistema de Información sanitario.

SKOS: Sistema simple de organización del conocimiento. Del inglés “Simple Knowledge Organization System”.

SNOMED-CT: Terminología clínica de nomenclatura médica sistematizada. Del inglés “Systematized Nomenclature of Medicine – Clinical Terms”.

SPARQL: Lenguaje de consulta y protocolo de SPARQL (acrónimo recursivo). Del inglés “SPARQL Protocol And RDF Query Language”.

STRIDE: Entorno de base de datos integrada de investigación traslacional de Stanford. Del inglés “Stanford Translational Research Integrated Database Environment”.

SWRL: Lenguaje de reglas de la web semántica. Del inglés “Semantic Web Rule Language”.

T2: Secuencia de IRM tipo T2.

T2W: Secuencia de IRM tipo T2 ponderada. Del inglés “T2 Weighted”

TAC: Tomografía Axial Computarizada.

TC: Tomografía Computarizada.

TLAP: Grupo consultor de librerías de plantillas. Del inglés “Template Library Advisory Panel”.

TI-RADS: Sistema de datos e informado de imágenes de. Del inglés “Imaging Reporting And Data System”.

URI: Identificador de recursos uniforme. Del inglés “Uniform Resource Identifier”.

US: Ultrasonidos. Del inglés “Ultra Sound”.

VOID: Vocabulario de conjuntos de datos interconectados. Del inglés “Vocabulary of Interlinked Datasets”.

W3C: Consorcio de la web mundial. Del inglés “World Wide Web Consortium”.

XML: Lenguaje de maquetación extensible. Del inglés “Extensible Markup Language”.

ZP: Zona Periférica.

Glosario

Aprendizaje automático: conjunto de técnicas que permiten a los sistemas informáticos predecir, clasificar, ordenar, tomar decisiones y, en general, extraer conocimientos de los datos sin necesidad de definir explícitamente las reglas para realizar esas tareas.

Arquetipo: representación formal de conceptos clínicos. Se vinculan a la terminología clínica. Son una combinación jerárquica de estructuras de modelos de referencia restringidas para ajustarse a la definición del concepto modelado.

Biomarcadores de imagen: parámetros cuantitativos anatómicos, fisiológicos, bioquímicos o moleculares detectables con métodos de procesamiento de imagen utilizados para establecer la presencia o gravedad de las enfermedades.

BioPortal: aplicación basada en la web para buscar, compartir, visualizar y analizar un gran repositorio de ontologías biomédicas, terminologías y anotaciones basadas en ontologías. Permite el mapeo entre términos, la exportación masiva de mapeos, la visualización de términos y relaciones dentro de las ontologías, las notas y la navegación de múltiples ontologías a través de pestañas.

BroadMatch: propiedad de mapeo de la ontología SKOS utilizada para establecer enlaces de mapeo jerárquicos entre conceptos SKOS en diferentes esquemas de concepto. Específicamente, se hace uso de BroadMach si el sujeto es taxonómicamente más amplio que el objeto.

CloseMatch: propiedad de mapeo de la ontología SKOS utilizada para establecer enlaces de mapeo. Específicamente, CloseMatch se emplea para enlazar dos conceptos que son suficientemente similares como para ser utilizados indistintamente en determinados contextos.

CLUSTER: entidad del modelo de referencia de las normas openEHR e ISO 13606 que sirve para organizar los elementos (componentes de tipo ELEMENT) de la HCE en grupos jerárquicos.

COMPOSITION: componente contenedor de información clínica de alto nivel del modelo de referencia de las normas openEHR e ISO 13606. Se corresponde con documentos clínicos

comúnmente utilizados, que incluyen la información emitida por un agente como resultado de un encuentro clínico.

DATA_VALUE: entidad del modelo de referencia de las normas openEHR e ISO 13606 que representa un valor atómico, con un tipo de dato básico y bien definido, dentro de un elemento (componente de tipo ELEMENT) de la HCE.

Datos enlazados: formato de datos estructurado que vincula los distintos datos distribuidos en Internet, y sus relaciones entre ellos, del mismo modo que se interconectan los enlaces de las páginas web.

Datos enlazados abiertos: repositorios públicos de datos enlazados a los que cualquiera puede acceder a cualquier dato que esté autorizado a ver, y también puede añadir a cualquier dato sin perturbar la fuente de datos original. Esto proporciona un entorno abierto en el que los datos pueden ser creados, conectados y consumidos a escala de Internet.

DrugBank: terminología con datos detallados de fármacos con información exhaustiva sobre objetivos y acciones de fármacos. Desde su primera versión en 2006, DrugBank se ha utilizado ampliamente para facilitar el descubrimiento de dianas farmacológicas, el diseño de fármacos, el acoplamiento o cribado de fármacos, la predicción del metabolismo de los fármacos, la predicción de las interacciones farmacológicas y la educación farmacéutica en general.

EHR_EXTRACT: entidad del modelo de referencia de las normas openEHR e ISO 13606 que representa la HCE de un paciente. Se trata del contenedor raíz de toda la HCE de un individuo.

ELEMENT: entidad del modelo de referencia de las normas openEHR e ISO 13606 que representa un único valor de información clínica. Corresponde con un nodo hoja de la estructura de la HCE.

ENTRY: entidad del modelo de referencia de las normas openEHR e ISO 13606 para registrar declaraciones clínicas. Incluye toda la información registrada en la HCE como resultado de una observación clínica, evaluación (interpretación clínica), instrucción (intención), o acción.

ExactMatch: propiedad de mapeo de la ontología SKOS utilizada para establecer enlaces de mapeo. Específicamente, ExactMatch se emplea para enlazar dos conceptos intercambiables en un amplio rango de contextos.

eXistDB: base de datos de código abierto NoSQL construida con tecnología XML. Se trata de una base de datos orientada a documentos XML de forma nativa, manipulable mediante el lenguaje de consulta XQuery.

FOLDER: componente organizador jerárquico de alto nivel del modelo de referencia de las normas openEHR e ISO 13606. Agrupa uno o más componentes COMPOSITION de acuerdo a diferentes criterios de ordenación.

Guías clínicas de radiología: pautas comunes para la evaluación de determinadas patologías, como las lesiones en próstata, pecho o hígado, entre otras áreas anatómicas.

Historia clínica electrónica (HCE): conjunto de documentos electrónicos donde se registra el historial médico de un paciente a lo largo de su vida. De acuerdo con la organización internacional de estandarización ISO, la HCE es un repositorio de información sobre el estado de salud de un sujeto de asistencia, en un formato computable, almacenado y transmitido de forma segura y accesible por múltiples usuarios autorizados. Dispone de un modelo lógico de información estandarizado o consensuado, que es independiente de los sistemas de información de apoyo a la asistencia sanitaria. Su principal objetivo es proporcionar una asistencia sanitaria continua, de calidad y eficiente, que contenga información retrospectiva, actual y prospectiva.

Informe radiológico estructurado (IRE): documento que integra el resultado final del proceso asistencial del radiólogo con el paciente. Contiene la información de forma estructurada, concisa y precisa, con campos separados y tabulados para que puedan ser visualizados y explotados posteriormente. Los IRE deben contener, al menos, los datos demográficos del paciente, el resumen de la información clínica relevante, la descripción de la técnica de exploración, la descripción detallada de los hallazgos y las conclusiones.

Interoperabilidad semántica: capacidad de dos o más sistemas informáticos para comprender el significado de la información que se transmite entre ellos, de forma automática y sin intervención humana.

Interoperabilidad sintáctica: capacidad de dos o más sistemas para transferir datos y documentos. Garantiza que la estructura de los documentos sea correcta.

Interoperabilidad técnica: capacidad de dos o más sistemas para comunicarse a nivel físico (cableado, enchufes, protocolos, etc.). Se ocupa de la transferencia de bytes.

ISO13606: estándar de HCE basado en el modelo dual, que describe la administración y almacenamiento de información sanitaria.

k_{ep} : en el contexto de los biomarcadores cuantitativos de IRM de próstata, se refiere al coeficiente de extracción que modela la vuelta del contraste extravasado al espacio vascular.

K^{trans} : en el contexto de los biomarcadores cuantitativos de IRM de próstata, se refiere a la constante de permeabilidad capilar, que mide el paso de contraste a través del endotelio desde el compartimento vascular al intersticio.

MappingRelation: propiedad de mapeo de la ontología SKOS utilizada para establecer enlaces de mapeo. Se trata de una propiedad abstracta cuyo significado indica que el sujeto está en una relación de mapeo con el objeto. De MappingRelation heredan otras relaciones como CloseMatch, broadMatch o NarrowMatch.

Modelo de conocimiento o dominio: representación conceptual de los elementos que forman parte de un área del conocimiento específica. Describe las entidades, sus atributos, sus roles y relaciones entre ellas, además de las restricciones inherentes al dominio. Las terminologías y ontologías médicas como SNOMED CT, DrugBank o NDFRT son ejemplos de modelos de dominio.

Modelo de información: representación formal y abstracta de entidades de un dominio y las relaciones entre ellas. La historia clínica electrónica y el informe radiológico estructurado son ejemplos de modelos de información.

Modelo de información clínica: estructura de datos estándar para el almacenamiento y la comunicación de la HCE, como openEHR, ISO 13606, HL7 CDA o DICOM-SR.

MongoDB: base de datos de código abierto NoSQL construida con tecnología BSON. Se trata de una base de datos orientada a documentos BSON de forma nativa, manipulable mediante el lenguaje de consulta MongoShell.

NarrowMatch: propiedad de mapeo de la ontología SKOS utilizada para establecer enlaces de mapeo jerárquicos entre conceptos SKOS en diferentes esquemas de concepto. Específicamente, se utiliza NarrowMach si el sujeto es taxonómicamente más acotado que el objeto.

NDFRT: National Drug File - Reference Terminology (Terminología de Referencia del Archivo Nacional de Fármacos) es la terminología utilizada por el organismo regulador de administración de alimentos y medicamentos del gobierno de Estados Unidos FDA (Food & Drug Administration) para codificar propiedades farmacológicas esenciales de los medicamentos: Mecanismo de acción, efecto fisiológico y clase estructural.

Nube de datos enlazados abiertos: grafo de conocimiento presentado como una web semántica de datos enlazados abiertos.

Ontología médica: descripción formal de los conceptos y relaciones en el dominio biomédico, donde se especifican los significados y las relaciones jerárquicas entre los términos y conceptos de dicho dominio.

OpenEHR: Estándar de HCE abierto, basado en el modelo dual de HCE, que describe la administración y almacenamiento de información sanitaria.

Radiómica: método de extracción de grandes volúmenes de características de las imágenes médicas mediante algoritmos de caracterización de datos. Estas características, denominadas radiómicas, tienen el potencial de descubrir patrones y características tumorales que no se aprecian a simple vista.

Radreport: repositorio de plantillas de IRE claras, coherentes y legibles tanto para las personas como para los ordenadores. Las plantillas han sido aportadas por radiólogos con experiencia en sus respectivas subespecialidades y revisadas por un panel internacional de expertos TLAP (Template Library Advisory Board). Cada plantilla de IRE está diseñada para permitir a los radiólogos captar las observaciones clínicas esenciales utilizando la terminología, las medidas, los parámetros técnicos, las anotaciones y las imágenes clave adecuadas. Las plantillas se ofrecen de forma gratuita y no están sujetas a restricciones de licencia para su uso. Se basan en las mejores prácticas y en las normas técnicas establecidas. Mejoran la comunicación entre los radiólogos y los especialistas destinatarios del informe.

Ayudan a las consultas de radiología a cumplir los criterios de acreditación y a obtener incentivos de pago por rendimiento.

RelatedMatch: propiedad de mapeo de la ontología SKOS, utilizada para establecer enlaces asociativos de mapeo entre dos conceptos SKOS en diferentes esquemas de concepto.

RECIST: método estándar que se usa para medir el modo en que un paciente de cáncer responde al tratamiento. Permite determinar si los tumores se reducen, permanecen igual o se agrandan.

Script: extracto de código en lenguaje programación que conforma un programa sencillo. Utilizados habitualmente para prototipar software, automatizar tareas repetitivas o realizar procesamiento por lotes.

SECTION: componente organizador jerárquico del modelo de referencia de las normas openEHR e ISO 13606, que proporciona una estructura lógica y de navegación para los datos dentro de un componente COMPOSITION. Refleja el flujo de recogida de información durante una sesión clínica de forma estructurada para facilitar su interpretación.

Sistema de ayuda a la decisión clínica (SADC): sistemas diseñados para ayudar a la toma de decisiones clínicas, en los que las características individuales de los pacientes se utilizan para generar recomendaciones que son presentadas a los clínicos para su consideración.

SKOS: Simple Knowledge Organization System (Sistema simple de organización del conocimiento), es una ontología común para compartir y enlazar sistemas de organización de conocimiento vía web.

SLOT: restricción del modelo de referencia definido en las normas openEHR e ISO 13606, que sirve para referenciar a un conjunto de posibles arquetipos válidos para definir una entidad dada de un arquetipo.

SNOMED CT: Systematized Nomenclature of Medicine - Clinical Terms (Nomenclatura Sistemática de Términos Clínicos) es la terminología clínica más completa, multilingüe y codificada desarrollada en el mundo. Está estructurada como un grafo dirigido y acíclico. Es, por tanto, una ontología médica.

Terminología médica: conjunto de términos que representan los conceptos en el campo específico de la medicina.

T-Rex: editor de plantillas de informe estructurado de la RSNA, que proporciona una interfaz gráfica para ensamblar componentes de plantillas o modificar una plantilla existente para crear una nueva.

v_e: en el contexto de los biomarcadores cuantitativos de IRM de próstata, se refiere a la fracción de volumen del espacio extracelular ocupada por el tumor.

Web semántica: extensión de la web mediante estándares establecidos por el consorcio de la web mundial W3C. El objetivo de la web semántica es hacer que los datos de Internet sean legibles por las máquinas. Para permitir la codificación semántica de los datos, se emplean tecnologías como RDF y OWL. Estas tecnologías permiten describir conceptos, relaciones entre entidades y su categorización, lo que ofrece ventajas como el razonamiento sobre los datos y el funcionamiento con fuentes de datos heterogéneas.

XQuery: lenguaje de consulta diseñado para colecciones de datos XML. Semánticamente se asemeja a SQL, aunque incluye capacidades adicionales de programación.

Resumen

La explotación de datos de salud ha demostrado ser de creciente interés en la comunidad científica, especialmente para la creación y uso de sistemas de ayuda a la decisión clínica (SADC). Para abordar este problema, tradicionalmente se ha investigado por separado en materia de modelos de información, modelos de dominio y SADC. En lo que se refiere a modelos de información, las propuestas presentan limitaciones semánticas y no tienen en cuenta la interacción con los modelos de dominio, que pretenden proporcionar una comprensión formal y compartida del conocimiento clínico, ni con los SADC, cuya finalidad es proporcionar apoyo a la toma de decisión clínica a partir de la historia clínica electrónica (HCE). La finalidad de esta tesis se enmarca dentro del objetivo general de enriquecer sistemas de HCE con resultados de SADC y modelos de dominio representados mediante datos enlazados abiertos. Para ello, se investiga la combinación y explotación conjunta de las tecnologías más avanzadas para modelos de información, modelos de dominio y SADC.

La principal contribución de esta tesis es el desarrollo de metodologías y herramientas para enriquecer la HCE con resultados de SADC y datos enlazados abiertos. Las contribuciones específicas son las siguientes:

- Definición conceptual y metodológica de la HCE aumentada con información potencialmente relevante de la web semántica.
- Definición conceptual y metodológica del informe radiológico estructurado (IRE), enriquecido con resultados de SADC basados en reglas, visión por computación y modelos de aprendizaje automático.
- Caso de uso de HCE aumentada, consistente en enriquecer la HCE resumida del Sistema Nacional de Salud de España con datos enlazados abiertos sobre interacciones farmacológicas y tratamientos recomendados para los episodios activos del paciente.
- Sistema de IRE enriquecido con resultados de SADC. Incluye el desarrollo de plantillas de IRE y mecanismos para el enriquecimiento de las mismas con resultados de SADC basados en reglas, cuantificación de imagen médica y redes neuronales.

Nuestro objetivo es mejorar el grado de interoperabilidad en las integraciones de sistemas de HCE con SADC y datos enlazados abiertos, mediante estrategias basadas en los tres pilares de la interoperabilidad semántica: modelos de información, de arquetipos y de dominio. Esto tiene el potencial de repercutir positivamente sobre la salud y el cuidado del paciente, especialmente en el paradigma de la medicina personalizada.

Resum

L'explotació de dades de salut ha demostrat ser de creixent interès en la comunitat científica, especialment per a la creació i ús de sistemes d'ajuda a la decisió clínica (SADC). Per a abordar aquest problema, tradicionalment s'ha investigat per separat en matèria de models d'informació, models de domini i SADC. Pel que fa a models d'informació, les propostes presenten limitacions semàntiques i no tenen en compte la interacció amb els models de domini, que pretenen proporcionar una comprensió formal i compartida del coneixement clínic, ni amb els SADC, la finalitat dels quals és proporcionar suport a la presa de decisió clínica partint de la història clínica electrònica (HCE). La finalitat d'aquesta tesi s'emmarca dins de l'objectiu general d'enriquir sistemes de HCE amb resultats de SADC i models de domini representats mitjançant dades enllaçades obertes. Per a això, s'investiga la combinació i explotació conjunta de les tecnologies més avançades per a models d'informació, models de domini i SADC.

La principal contribució d'aquesta tesi és el desenvolupament de metodologies i eines per a enriquir la HCE amb resultats de SADC i dades enllaçades obertes. Les contribucions específiques són les següents:

- Definició conceptual i metodològica de la HCE augmentada amb informació potencialment rellevant de la web semàntica.
- Definició conceptual i metodològica de l'informe radiològic estructurat (IRE), enriquit amb resultats de SADC basats en regles, visió per computació i models d'aprenentatge automàtic.
- Cas d'ús de HCE augmentada, consistent a enriquir la HCE resumida del Sistema Nacional de Salut d'Espanya amb dades enllaçades obertes sobre interaccions farmacològiques i tractaments recomanats per als episodis actius del pacient.
- Sistema de IRE enriquit amb resultats de SADC. Inclou el desenvolupament de plantilles de IRE i mecanismes per a l'enriquiment de les mateixes amb resultats de SADC basats en regles, quantificació d'imatge mèdica i xarxes neuronals.

El nostre objectiu és millorar el grau d'interoperabilitat en les integracions de sistemes de HCE amb SADC i dades enllaçades obertes, mitjançant estratègies basades en els tres pilars de la interoperabilitat semàntica: models d'informació, d'arquetips i de domini. Això té el potencial de repercutir positivament sobre la salut i la cura del pacient, especialment en el paradigma de la medicina personalitzada.

Summary

The exploitation of health data has proven to be of increasing interest in the scientific community, especially for the creation and use of clinical decision support systems (CDSS). To address this problem, separate research has traditionally been done on information models, domain models and CDSS. Regarding information models, the proposals present semantic limitations and do not consider the interaction with domain models, which aim to provide a formal and shared understanding of clinical knowledge, nor with CDSS, whose purpose is to provide clinical decision support from the electronic health record (EHR). The aim of this thesis is framed within the general goal of enriching EHR systems with SADC results and domain models represented by open linked data. For this purpose, the combination and joint exploitation of state-of-the-art technologies for information models, domain models and SADC is investigated.

The main contribution of this thesis is the development of methodologies and tools to enrich EHR with SADC results and open linked data. Specific contributions are:

- Conceptual and methodological definition of EHR augmented with potentially relevant information from the semantic web.
- Conceptual and methodological definition of the structured radiology report (SRR), enriched with CDSS results based on logical rules, computer vision and machine learning models.
- Augmented EHR use case, consisting of enriching the summarized EHR of the Spanish National Health System with linked open data on pharmacological interactions and recommended treatments for active patient episodes.
- SRR system enriched with CDSS results. Includes the development of SRR templates and mechanisms for enriching them with SADC results based on logical rules, medical image quantification and neural networks.

Our goal is to improve the degree of interoperability in EHR system integrations with CDSS results and linked open data, through strategies based on the three pillars of semantic interoperability: information, archetype and domain models. This has the potential to positively impact health and patient care, especially in the personalized medicine paradigm.

Capítulo 1. Introducción

1.1. Motivación

La historia clínica electrónica (HCE) constituye una pieza clave de la asistencia sanitaria del paciente, especialmente porque el paradigma de los servicios sanitarios ha ido cambiando hacia un escenario centrado en el paciente [1]. La integridad, exactitud e interoperabilidad de la HCE tiene un impacto directo en los resultados de la atención sanitaria del paciente [2], pero también en la explotación y reuso de los datos de salud, que ha demostrado ser de creciente interés en la comunidad científica, por ejemplo, para la creación y uso de sistemas de ayuda a la decisión clínica (SADC) [3]–[14]. Para abordar este problema, tradicionalmente se ha investigado por separado en materia de modelos de información, modelos de dominio y SADC. En lo que se refiere a modelos de información, las propuestas presentan limitaciones semánticas [15] y no tienen en cuenta ni la interacción con los modelos de dominio, que pretenden proporcionar una comprensión formal y compartida del conocimiento clínico, ni con los SADC, cuya finalidad es proporcionar apoyo a la toma de decisión clínica a partir de la HCE. El sector sanitario cada vez produce y consume mayor cantidad de datos e información. El intercambio de información es la norma y no la excepción, aunque un intercambio de datos semánticamente interoperable por completo está lejos de ser conseguido. El cambio predecible hacia la medicina personalizada producirá un aumento drástico en el tamaño y la complejidad de los sistemas de HCE, lo que a su vez afectará a la integración y comunicación de los datos clínicos. La consecución de un alto nivel de interoperabilidad semántica es uno de los retos más importantes en la actualidad [16]–[22]. La interoperabilidad semántica es vital para asegurar la coherencia global del significado de los datos clínicos, un requisito básico para permitir una mejor atención sanitaria, así como para su reuso en investigación clínica.

Desde el punto de vista asistencial, idealmente los sistemas de HCE deben interactuar con los SADC. Esto contribuye a la automatización de los procesos de atención con el consiguiente ahorro de tiempo y recursos, y al mismo tiempo el aumento de la seguridad del paciente. Independientemente del tipo de tarea, los SADC deben integrarse con los sistemas de HCE utilizados rutinariamente por los médicos. La cooperación entre modelos de información, dominio y SADC se considera como esencial en la literatura, pero la comunidad científica no se ha ocupado de explotar este enfoque hasta la fecha. El hecho de que sea un objetivo a largo plazo de la red europea de excelencia SemanticHealthNet demuestra su relevancia. Dicha red establece los modelos de información (modelos de referencia y de arquetipos), y los modelos de dominio (terminologías y ontologías) como los pilares básicos de los sistemas semánticamente interoperables [23]. Por otra parte, los SADC mejorarán si

son capaces de utilizar más información y, sobre todo, información de más calidad. En este contexto, las iniciativas que promueven la compartición de datos en forma de datos enlazados están recibiendo un gran impulso. Los datos enlazados deben ser tomados en cuenta como posibles fuentes de información. Un número creciente de organizaciones biomédicas publican conjuntos de datos de alta calidad sobre enfermedades, medicamentos, interacciones, ensayos clínicos, etc., en formato de datos enlazados. Estos pueden ser explotados para apoyar los procesos de toma de decisiones.

En esta tesis se aborda el problema de enriquecer la HCE con información de fuentes de conocimiento heterogéneas y resultados de SADC, mediante soluciones que integran con un alto grado de interoperabilidad semántica sistemas de información sanitarios, fuentes de conocimiento de la web semántica y resultados de SADC. Para ello, se hace hincapié en el uso de los pilares básicos de los sistemas semánticamente interoperables descritos con anterioridad. La interoperabilidad semántica es un reto, que ni en el plano académico ni en la práctica clínica ha sido solventado hasta la fecha en su totalidad. Como se explica más adelante en el capítulo 2, existen multitud de trabajos en la literatura dedicados a la interoperabilidad semántica de la información de salud que, si bien no tienen una aplicación práctica extendida hoy en día, es probable que la tengan en el futuro.

Asociar conocimiento externo potencialmente relevante al caso clínico, explotando las sinergias de la información de salud con otras fuentes de conocimiento heterogéneas, puede resultar fundamental en determinados casos en los que dicho conocimiento adicional contribuye a realizar un diagnóstico acertado o a identificar un tratamiento adecuado. Aumentar la HCE con información adicional sobre interacciones medicamentosas, tratamientos recomendados para episodios del paciente, o recomendaciones médicas personalizadas de acuerdo con el genotipo del paciente, son algunos ejemplos de HCE aumentada que pueden impulsar el paradigma de la medicina personalizada centrada en el paciente, enriqueciendo su HCE con información contextual potencialmente relevante para su caso clínico. Concretamente, en esta tesis se investiga un enfoque nuevo para facilitar la interoperabilidad semántica de información de salud con fuentes de conocimiento heterogéneas, que no ha sido abordado previamente en la literatura. Además, describe la investigación realizada sobre cómo resolver la problemática de construir sistemas de HCE interoperables con resultados de SADC, y cómo dichos sistemas pueden llevarse a la práctica clínica. La literatura pone de manifiesto la necesidad de sistemas de información sanitarios interoperables con fuentes de conocimiento de dominio clínico y resultados de SADC [1]-[23], que justifican el esfuerzo realizado en esta tesis. La revisión bibliográfica presentada en el capítulo 2 refuerza este argumento, identificando las limitaciones actuales de este campo de creciente interés en la comunidad científica. Además, en capítulos posteriores se presentan discusiones sobre cómo los resultados de esta tesis contribuyen a mejorar el estado de la

técnica, comparando los resultados alcanzados con las soluciones actuales disponibles en la literatura.

1.2. Hipótesis, preguntas de investigación y objetivos

1.2.1. Hipótesis

La finalidad de esta tesis se enmarca dentro del objetivo general de enriquecer sistemas de HCE con SADC y modelos de dominio representados mediante datos enlazados abiertos. Para ello, se plantea la hipótesis de investigación de que la combinación y explotación conjunta de las tecnologías más avanzadas para los modelos de información, de dominio y SADC pueden facilitar la interoperabilidad entre los sistemas de HCE, los SADC y fuentes de datos enlazados abiertos. Nótese que el informe radiológico estructurado (IRE) forma parte de la HCE, y tanto la HCE como el IRE representan modelos de información clínica adecuados para los objetivos de estudio de esta tesis. El desarrollo de este trabajo requiere la combinación de conocimientos y técnicas avanzadas en modelos de información para sistemas HCE, terminologías y ontologías clínicas de la web semántica, y SADC. Concretamente, en esta tesis se investiga la combinación de herramientas, metodologías y normas con impacto en la interoperabilidad de los sistemas de HCE con SADC y fuentes de datos enlazados abiertos, concretamente: (1) Arquitecturas de HCE basadas en arquetipos y plantillas, para los modelos de información; (2) Sistemas basados en reglas, visión por computación y modelos de aprendizaje automático para los SADC; y (3) Terminologías, ontologías y datos enlazados del área de la web semántica, para el modelo de dominio. Nuestra aproximación persigue reutilizar al máximo las tecnologías existentes, preferiblemente disponibles como software libre, para fomentar la transferencia tecnológica en el marco de esta tesis. Además, el uso de los tres pilares básicos de la interoperabilidad semántica: modelos de información, de arquetipos y de dominio, contribuye a mejorar la interoperabilidad de los desarrollos de la tesis con los sistemas de HCE, así como la representación de cualquier tipo de dato clínico. En la base del desarrollo están los estándares de HCE ISO13606, imagen médica DICOM, terminologías u ontologías como SNOMED CT [24], DrugBank [25] y NDFRT [26], y herramientas normalizadas de la web semántica para tratar con fuentes de datos enlazados abiertos.

1.2.2. Preguntas de investigación y objetivos

Se plantean las siguientes preguntas de investigación con el objetivo de confirmar la hipótesis:

Pr1	¿Es posible aumentar la HCE, normalizada con arquetipos ISO13606, con contenidos potencialmente relevantes de fuentes de datos enlazados abiertos?
Pr2	¿Es posible enriquecer el IRE, normalizado con plantillas DICOM-SR, con información resultante de SADC basados en reglas, visión por computación y modelos de aprendizaje automático?

Para responder a estas preguntas de investigación, se plantean tres objetivos principales, centrados en la combinación y explotación conjunta de las tecnologías más avanzadas para modelos de información, modelos de dominio y SADC. Específicamente, se investigan soluciones de interoperabilidad entre sistemas de HCE, SADC y datos enlazados abiertos (Objetivo O1), se desarrolla del concepto de HCE aumentada con información de datos enlazados abiertos (Objetivo 2) y el concepto de IRE enriquecido con resultados de SADC (Objetivo O3), dando lugar a los objetivos y subobjetivos listados a continuación:

O1	Diseñar metodologías para enriquecer sistemas de HCE con información de SADC y datos enlazados abiertos.
O1.1	Definir transformaciones para la integración de información de SADC en la HCE.
O1.2	Definir transformaciones para la integración de información de dominio en la HCE.
O2	Desarrollar el concepto de HCE aumentada y mecanismos para la creación, uso, persistencia y consulta de la HCE aumentada.
O2.1	Definir conceptualmente la HCE aumentada.
O2.2	Aumentar la HCE con datos enlazados abiertos.
O2.3	Proporcionar mecanismos de persistencia y consulta de la HCE aumentada.
O3	Desarrollar el concepto de IRE enriquecido con resultados de SADC y mecanismos para la creación, uso, persistencia y consulta del IRE enriquecido.
O3.1	Definir conceptualmente el IRE enriquecido.
O3.2	Enriquecer el IRE con resultados de SADC.
O3.3	Proporcionar mecanismos de persistencia y consulta del IRE enriquecido.

1.3. Contribuciones

A continuación, se describen las contribuciones principales de esta tesis. Cada contribución está relacionada con uno o más objetivos, e incluye referencias a publicaciones en congresos y/o revistas científicas donde la contribución ha sido publicada (publicaciones P1 a P3 reseñadas en esta sección):

C1.	Definición conceptual y metodológica de la HCE aumentada con información potencialmente relevante de la web semántica.
	Contribuye a los objetivos de investigación O1, O1.2, O2, O2.1, O2.2
	Publicado en P1
	La HCE de un paciente se puede complementar con información adicional potencialmente relevante, que puede ser útil para usos secundarios de la HCE o los SADC. Dicha información puede obtenerse de fuentes de datos enlazados abiertos. Esta tesis define la HCE aumentada como una HCE convencional complementada con contenidos adicionales procedentes de fuentes de datos enlazados abiertos. Información y/o conocimiento de potencial interés para su inclusión en la HCE es susceptible de formar parte de una HCE aumentada. Garantizar una integración completa de la HCE con una fuente externa requiere obtener una representación común de contenidos en distintos formatos. Con el objetivo de ser ampliamente aplicable, se proporciona una definición metodológica formal basada en estándares de HCE y datos enlazados abiertos de la web semántica. El proceso parte de instancias de HCE normalizadas de acuerdo a arquetipos de la norma ISO 13606, y contenido de aumento expresado en datos enlazados abiertos. Mediante especificaciones de mapeo, se establecen las relaciones existentes entre ambas fuentes para garantizar la coherencia y dominio semántico de la información. Posteriormente, a partir de las especificaciones de mapeos se implementan transformaciones para convertir instancias de datos de diferentes fuentes.

C2.	Definición conceptual y metodológica del IRE enriquecido con información resultante de SADC basados en reglas, visión por computación y modelos de aprendizaje automático.
	Contribuye a los objetivos de investigación O1, O1.1, O3, O3.1, O3.2.
	Publicado en P2 y P3.
	Las ventajas del IRE respecto al informe radiológico, tradicionalmente basado en dictado en radiología, se describen ampliamente a lo largo de esta tesis. Además, se identifican las limitaciones actuales para la implantación completa del IRE en los servicios de radiología, y se propone el IRE enriquecido como una solución para mitigar dichas limitaciones. Conceptualmente, el IRE enriquecido consiste en un IRE convencional complementado con resultados de distintos tipos de SADC. Dotado, además, de funcionalidades para generar extractos de lenguaje natural a partir de los contenidos del informe, e implementar validadores de campos y reglas lógicas para asistir el proceso de informado. Estas cualidades contribuyen a minimizar la tasa de errores en el informado, poner los resultados de SADC a disposición del radiólogo durante el informado, facilitar la transición del informado tradicional al IRE, homogeneizar el informado radiológico y adecuar el contenido del informe al público objetivo.

C3.	Caso de uso de HCE aumentada, consistente en enriquecer la HCE resumida del Sistema Nacional de Salud de España con datos enlazados abiertos sobre interacciones farmacológicas y tratamientos recomendados para los episodios activos del paciente.
	Contribuye a los objetivos de investigación O1, O1.2, O2, O2.2, O2.3.
	Publicado en P1
	Implementar una HCE aumentada requiere de la definición de arquetipos, transformaciones entre esquemas, y una fuente de información externa con el contenido de aumento. En este trabajo, presentamos un caso de uso de HCE aumentada que permite enriquecer extractos de HCE normalizados con diferentes aumentos de HCE. Para el caso de uso, se han diseñado dos aumentos para la HCE resumida del Sistema Nacional de Salud de España: interacciones farmacológicas para la medicación del paciente y tratamientos recomendados para

	<p>los episodios activos del paciente. A partir de un extracto de HCE resumida, el sistema consulta las fuentes externas de datos enlazados abiertos y ejecuta las transformaciones necesarias para generar una instancia de HCE resumida aumentada. Aumentar una HCE con esta implementación es automático. Sin embargo, el modelado de los recursos necesarios para definir un aumento de HCE requiere de ciertos pasos manuales. Para asistir el proceso, se presentan algoritmos que automatizan algunos de esos pasos.</p>
--	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

C4	<p>Sistema de IRE enriquecido con resultados de SADC. Incluye el desarrollo de plantillas de IRE y mecanismos para el enriquecimiento de las mismas con resultados de SADC basados en reglas, cuantificación de imagen médica y redes neuronales.</p>
	<p>Contribuye a los objetivos de investigación O1, O1.1, O3, O3.2, O3.3.</p>
	<p>Publicado en P2 y P3</p>
	<p>Presentamos un sistema de IRE que integra resultados de distintos tipos de SADC: (I) basados en reglas para calcular puntuaciones cualitativas de acuerdo a guías clínicas de radiología (LI-RADS, BI-RADS o PI-RADS). (II) Basados en cuantificación de imagen médica para la obtención de parámetros cuantitativos relevantes para el diagnóstico, y (III) en redes neuronales para la detección de lesiones y regiones de interés. La herramienta gestiona la creación, especialización, versionado y borrado de plantillas de formulario de IRE de acuerdo al perfil de integración IHE MRRT para la gestión de plantillas de IRE. Además, las plantillas de IRE de nuestro sistema son compatibles con las plantillas de la iniciativa Radreport de RSNA. A partir de las valoraciones cualitativas del radiólogo, se calculan automáticamente las puntuaciones finales de los sistemas de puntuación cualitativos. Los resultados cuantitativos de SADC se embeben en los formularios de IRE mediante objetos JSON y marcadores de posición. Al finalizar el informe, el sistema es capaz de generar el contenido en lenguaje natural y en diferentes idiomas, adaptando el IRE a la audiencia destinataria del informe. El sistema permite exportar los IRE en formatos normalizados (HL7 y DICOM SR), facilitando la interoperabilidad del IRE con los sistemas de información sanitarios.</p>

Publicaciones científicas

El trabajo presentado en esta tesis se extiende a lo largo de ocho años de investigación, desarrollo y trabajo en empresas. Las siguientes publicaciones en revistas y contribuciones a conferencias describen los resultados alcanzados y reflejan el trabajo realizado en proyectos de investigación específicos y soluciones comerciales, que requerían en mayor o menor grado soluciones de interoperabilidad de modelos de información (arquetipos de HCE y plantillas de IRE), con modelos de dominio (terminologías y ontologías) y SADC.

Publicaciones en revistas	
P1	A. Mañas-García, J. A. Maldonado, M. Marcos, D. Boscá, y M. Robles, «Augmented EHR: Enrichment of EHR with Contents from Semantic Web Sources», Appl. Sci., vol. 11, n.o 9, Art. n.o 9, ene. 2021, doi: 10.3390/app11093978.
P2	A. Mañas-García, I. González-Valverde, E. Camacho-Ramos, A. Alberich-Bayarri, J. A. Maldonado, M. Marcos, y M. Robles, «Radiological Structured Report Integrated with Quantitative Imaging Biomarkers and Qualitative Scoring Systems», J. Digit. Imaging, feb. 2022, doi: 10.1007/s10278-022-00589-9.
Contribuciones a congresos	
P3	A. Mañas-García, E. Camacho-Ramos, I. Gonzalez-Valverde, F. Bacha-Villamide, A. Alberich-Bayarri, y L. Martí-Bonmatí, «Implementation of an interactive radiological structured report management system with AI annotation capabilities», presentado en European Society of Medical Imaging Informatics (EuSoMII) Annual Meeting 2019. Medical Imaging Informatics. AI, Clinical Applications and more., 2019.

Software

Durante la elaboración de esta tesis se han desarrollado los siguientes casos de uso y sistemas software:

Caso de uso de HCE aumentada

Aplicación web que, a partir de un extracto de HCE y un identificador de aumento de HCE, ejecuta las transformaciones necesarias para construir la versión aumentada del extracto de HCE proporcionado: (I) extracción de las entidades de interés del extracto de HCE. (II) Consulta de servicio de mapeo terminológico para identificar las entidades de interés expresadas en términos de la fuente externa. (III) Consulta de la fuente externa para obtener el contenido de aumento. (IV) Transformación del contenido de aumento a formato de HCE normalizada. (V) Integración del contenido de aumento normalizado con el extracto inicial de HCE. Además, se incluye un repositorio para persistir aumentos de HCE, contemplando que para cada arquetipo de HCE se pueden definir diferentes aumentos de HCE, y facilitando al usuario escoger el/los aumento(s) más adecuado(s) para cada extracto de HCE. También dispone de una interfaz para gestionar los aumentos de HCE definidos (consultar, crear, borrar o modificar aumentos de HCE) y un sistema de versionado para llevar un seguimiento de la evolución de la HCE aumentada.

Sistema de IRE enriquecido

Plataforma web para la gestión y el uso del IRE enriquecido. Se trata de un sistema que permite importar plantillas de IRE como módulos independientes. Cada plantilla contiene las especificaciones de un formulario de IRE para una determinada patología o área anatómica. El formulario recoge la información relevante del caso que le proporciona el radiólogo, homogeneizando el informado de cada patología y/o área anatómica. El sistema permite generar extractos del informe en lenguaje natural, adaptando el contenido del informe al público objetivo. Las plantillas del informe, están dotadas de mecanismos de reglas para modelar SADC como las calculadoras de puntuación PI-RADS, LI-RADS o BI-RADS. Dichos mecanismos se aplican también para restringir campos del formulario y evitar errores durante el informado. Por último, las plantillas cuentan con funcionalidades para integrar resultados de SADC externos, calculados en procesos complejos al margen de la plantilla, tales como cuantificación de imagen médica o modelos de aprendizaje automático para detección de lesiones y regiones de interés.

Proyectos de investigación

Parte de esta tesis se enmarca en el contexto de un proyecto de I+D del campo de la interoperabilidad de la información de salud. La experiencia obtenida de este proyecto, ha permitido identificar las principales limitaciones y necesidades de interoperabilidad de sistemas de HCE, tanto con SADC como con otras fuentes de información. Además, el resultado de la investigación realizada en este proyecto ha llevado a definir la HCE aumentada, a diseñar una metodología de creación y uso de HCE aumentada, así como a proporcionar mecanismos de persistencia y consulta de HCE aumentada.

Proyecto	Modelos de información y conocimiento clínicos para enlazar los sistemas de historia clínica electrónica y de ayuda a la decisión clínica (TIN2014-53749-C2-1-R)
Entidad	Universitat Politècnica de València y Universitat Jaume I

Colaboraciones con empresas

La colaboración con la compañía Quibim SL, especializada en SADC basados en biomarcadores de imagen, ha proporcionado la experiencia y herramientas necesarias para diseñar, implementar y llevar a la práctica clínica el sistema de IRE enriquecido. El IRE forma parte de la HCE, y la solución de IRE enriquecido dota a la HCE de interoperabilidad semántica parcial con SADC basados en reglas, parámetros cuantitativos de biomarcadores de imagen y detección de lesiones mediante modelos de inferencia.

Otras contribuciones

Durante el desarrollo de la tesis, el autor ha participado en la preparación de material y en la impartición de clases tanto teóricas como prácticas en asignaturas, proyectos fin de grado y ponencias listadas a continuación.

- Máster interuniversitario de ingeniería biomédica de la Universidad de Valencia y la Universidad Politècnica de Valencia. Asignaturas:
 - “Calidad de datos e interoperabilidad”. Aprendizaje y práctica de estándares de HCE, integración de HCE con SADC mediante arquetipos y técnicas de curación de datos basadas en transformaciones entre modelos de datos estándares. (2016 – 2020).
 - “Sistemas de Información médica avanzados”. Aprendizaje de metodologías de diseño y gestión de sistemas de información “Soft system methodology” (2014 y 2015).

- Grado de ingeniería biomédica de la Universidad Politécnica de Valencia. Asignaturas y proyectos:
 - Proyecto final de grado tutorizado: “Modelado de recursos HL7 FHIR mediante arquetipos del estándar ISO 13606” (2016 - 2017).
 - Asignatura “Imágenes biomédicas”. Clases prácticas de técnicas de procesado de imagen y visión por computación para la extracción de parámetros cuantitativos de imagen médica y delimitación de regiones de interés (2016 - 2017).
 - Asignatura “Sistemas de información y telemedicina II”. Impartición teórica de interoperabilidad semántica de la HCE, normalización de la HCE, modelo de HCE dual, modelos de HCE no duales, modelo de arquetipos, casos de aplicación, ontologías y web semántica en biomedicina (2015 - 2017).
- Máster en gestión sanitaria: dirección de hospitales y servicios de salud, del centro de investigación en economía y gestión de la salud (CIEGS). Universidad Politécnica de Valencia.
 - Charla sobre interoperabilidad e interconectividad de sistemas de información hospitalarios (2017).

Además, el autor ha colaborado en contribuciones a congresos científicos relacionadas con el contenido de esta tesis:

P4	B. Martínez-Salvador, M. Marcos, A. Mañas-García, J. A. Maldonado, y M. Robles, «Using SNOMED CT Expression Constraints to Bridge the Gap Between Clinical Decision-Support Systems and Electronic Health Records.», Stud. Health Technol. Inform., vol. 228, pp. 504-8, 2016
P5	A. Mañas-García, R. S. Requena, A. Alberich-Bayarri, G. García-Martí, L. A. Tortajada, E. R. Martínez, y L. Martí-Bonmatí, «Imaging biomarkers processing platform», en IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), jun. 2014, pp. 416-419. doi: 10.1109/BHI.2014.6864391.
P6	J. C. Crespo, A. Mañas-García, D. B. Tomás, y M. Robles Viejo, «Metodología para el modelado de Recursos HL7 FHIR mediante arquetipos del estándar ISO13606», en CASEIB 2016 XXXIV Congreso Anual de la Sociedad Española de Ingeniería Biomédica: Libro de actas, 2016, ISBN 978-84-9048-531-6, págs. 147-150, 2016, pp. 147-150.

P7	R. Sanz-Requena, A. Mañas-García, M. Robles Viejo, y L. Martí-Bonmatí, «Web application for the integration of bone age calculation in RIS/PACS», ECR 2013 EPOS, 7 de marzo de 2013. https://epos.myesr.org/poster/esr/ecr2013/C-0493 (accedido 20 de febrero de 2022).
P8	A. Mañas-García, R. Sanz-Requena, y M. Robles Viejo, «Aplicación Web para la integración del cálculo de la edad ósea con los sistemas de información del entorno hospitalario», presentado en XXX Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2012), San Sebastián, España, 2012.

1.4. Estructura de la tesis

La tesis está estructurada en seis capítulos. La figura 1 muestra el orden de los mismos y cómo están relacionados entre sí. El capítulo 1, “Introducción”, presenta una visión general del trabajo de investigación realizado y la motivación que ha llevado a su elaboración. Establece la hipótesis, preguntas de investigación para comprobar la hipótesis y objetivos para responder a las preguntas de investigación. Además, este capítulo describe las contribuciones, publicaciones científicas en congresos y revistas indexadas JCR, proyectos de investigación, colaboraciones y docencia impartida por el autor durante el desarrollo de esta tesis. El capítulo 2, “Estado del arte”, incluye una revisión de la literatura científica centrada en soluciones de interoperabilidad entre sistemas de HCE, SADC y otras fuentes de información heterogéneas, contemplando la HCE desde una perspectiva amplia que incluye al IRE. El capítulo 3, “Historia clínica electrónica aumentada con datos enlazados abiertos”, proporciona una definición de la HCE aumentada, una metodología formal para su diseño e implementación, y un caso de uso de HCE aumentada en el que la HCE resumida del Sistema Nacional de Salud de España es enriquecida con información de interacciones farmacológicas de la medicación activa del paciente, e información sobre posibles tratamientos para los episodios activos del mismo. El capítulo 4, “Informe radiológico estructurado enriquecido con sistemas de ayuda a la decisión clínica”, introduce el IRE enriquecido como una solución para mitigar algunas de las limitaciones del IRE, integrando resultados de distintos tipos de SADC con el informe radiológico. Además de una definición de IRE enriquecido, también describe una metodología para la obtención del mismo, y un sistema que gestiona su ciclo de vida e integración con distintos tipos de SADC. En el capítulo 5, “Discusión”, se discuten los resultados obtenidos en las contribuciones de esta tesis, se

compara con soluciones existentes en la literatura y se razona en qué aspectos nuestra solución contribuye a mejorar el estado del arte. Además, se plantean estrategias para combinar la HCE aumentada y el IRE enriquecido. Por último, el capítulo 6, “Conclusión y trabajo futuro”, aborda las conclusiones derivadas de esta tesis, valora la consecución de la hipótesis y los objetivos propuestos, e identifica posibles líneas de investigación como trabajo futuro. La figura 1 ilustra la relación entre los capítulos descritos y los anexos de este trabajo:

- Anexo I. Plantillas de informe radiológico estructurado
- Anexo II. Instituciones sanitarias y/o de I+D+i utilizando el sistema de informe radiológico estructurado

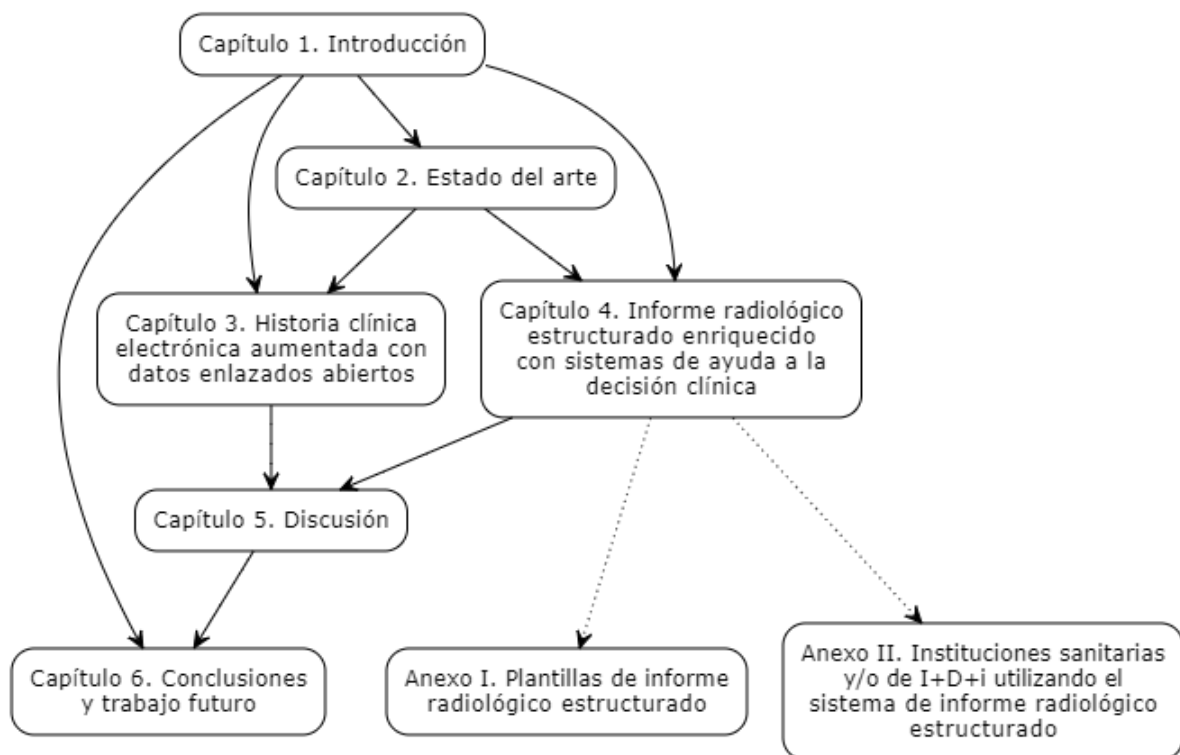


Figura 1. Estructura de la tesis. Orden y relaciones entre capítulos y anexos.

Por último, la tabla 1 muestra la relación entre los capítulos de la tesis y las preguntas de investigación, objetivos, contribuciones y publicaciones realizadas.

Capítulo	Preguntas de investigación	Objetivos	Contribuciones	Publicaciones
1. Introducción				
2. Estado del arte				
3. Historia clínica electrónica aumentada con fuentes de información heterogéneas	Pr1	O1, O1.2, O2, O2.1, O2.2, O2.3	C1, C3	P1
4. Informe radiológico estructurado enriquecido con sistemas de ayuda a la decisión clínica	Pr2	O1, O1.1, O3, O3.1, O3.2, O3.3	C2, C4	P2 y P3
5. Discusión				
6. Conclusiones y trabajo futuro				

Tabla 1. Estructura de la tesis y relación de capítulos con preguntas de investigación, objetivos y publicaciones.

Capítulo 2. Estado del arte

Esta tesis investiga el enriquecimiento de la HCE con resultados de SADC y datos enlazados abiertos LOD (Linked Open Data). Como punto de partida, se describen las razones que han llevado a escoger los SADC y los datos enlazados como fuentes de información para aumentar la HCE.

La evolución de la HCE ha motivado el desarrollo de diferentes tipos de SADC y metodologías para la integración de los mismos con los datos clínicos de la HCE. La explosión de datos sanitarios ha sido impulsada tanto por las fuerzas reguladoras como por la mayor disponibilidad tecnológica para apoyarla. Los datos de salud no solo se registran en repositorios digitales, sino que permiten identificar tendencias para su interpretación en la próxima visita al médico, o estar rápidamente disponibles para otros profesionales médicos en situaciones de emergencia [27]. La interoperabilidad de la información de salud contenida en las HCEs juega un papel crucial en la integración con SADC. Permite a los sistemas de HCE y los SADC intercambiar información de salud e interpretarla a nivel sintáctico y, en algunos casos, también a nivel semántico. En esta tesis, uno de los principales objetivos es el enriquecimiento de la HCE con SADC basados en biomarcadores de imagen, integrando sus resultados con una parte fundamental de la HCE como es el IRE.

En un esfuerzo por mejorar la accesibilidad de los datos y el conocimiento, la iniciativa de datos enlazados LD (Linked Data) propuso un conjunto bien definido de recomendaciones para exponer, compartir e integrar datos, información y conocimiento, utilizando tecnologías de la web semántica. Los datos enlazados abiertos LOD, son el resultado de la aplicación de los principios de LD a conjuntos de datos abiertos (es decir, datos que pueden utilizarse y distribuirse libremente). Existe una tendencia al uso de técnicas de datos LOD en ciencias de la salud, donde muchos estudios representan y gestionan su conocimiento mediante RDF (Resource Description Framework) y ontologías biomédicas [28]. Las tecnologías de la web semántica están teniendo un fuerte impacto en cómo se genera el conocimiento a partir de los datos biomédicos, haciendo que los datos estén cada vez más conectados y proporcionando una mejor descripción de su semántica. Otro objetivo principal de nuestro trabajo de investigación es la HCE aumentada, que consiste en enriquecer HCE normalizadas con contenido adicional, potencialmente relevante, procedente de fuentes de información de la web semántica.

Los sistemas de información sanitaria, la HCE, los estándares de HCE e imagen médica, el IRE, los SADC y la web semántica, son algunas de las herramientas más relevantes que han contribuido a la evolución de la HCE, y hacen posible que en esta investigación nos planteemos enriquecer la HCE con resultados de SADC e información procedente de fuentes de la web semántica. Las siguientes secciones son un repaso a la literatura sobre estas

herramientas clave (secciones 2.1 - 2.7) y otros trabajos relacionados con la interoperabilidad entre sistemas de HCE y SADC y/o herramientas de la web semántica (sección 2.8).

2.1. Sistemas de información sanitarios

Un sistema de información sanitario [29] es aquel sistema de información cuyo dominio principal es la sanidad. Gestiona colecciones de datos para la gestión clínica y la gestión administrativa. Se caracteriza por su enorme complejidad, tanto por su extensión como por tener que lidiar con modelos de datos muy ricos y cambiantes. Los sistemas de información sanitarios deben gestionar la integración y la interoperabilidad entre diversos subsistemas de información internos y externos. Las funciones básicas de un sistema de información sanitario incluyen la educación e investigación, adquisición y presentación de datos, administración de registros, comunicación e integración de datos, vigilancia y trazabilidad, almacenamiento y recuperación, análisis de datos y soporte en la toma de decisiones. Ejemplos destacables de sistemas de información sanitarios son: sistema de información poblacional, sistemas de información de radiología, sistemas de información de laboratorio, sistema de salud pública, farmacia, sistema de atención primaria, sistema de atención especializada, administración, y sistema de HCE. El trabajo de investigación de esta tesis se centra en dos de los sistemas de información sanitarios más importantes, concretamente los sistemas de información de radiología, y de HCE:

- Sistema de HCE: Gestiona la HCE. Mantiene un registro central de pacientes, y referencias a cada porción de información que confecciona la HCE de un paciente.
- Sistema RIS (Radiology Information System): Gestiona los informes radiológicos. Para cada paciente, registra sus informes radiológicos y referencias a las imágenes médicas. También se encarga de las peticiones de exámenes radiológicos, citación en radiología y gestión de resultados.
 - Sistema PACS (Picture Archiving and Communication System): Repositorio de imágenes médicas. Permite almacenar y consultar imágenes médicas adquiridas de los pacientes.

El RIS, a su vez, se integra con el hardware de adquisición de imágenes médicas de diferentes modalidades, como por ejemplo resonancia magnética, tomografía computarizada, o radiografía convencional, entre otros (modalidades de adquisición, en adelante).

La figura 2 presenta estos subsistemas de información sanitarios y los principales flujos de trabajo (FT) que involucran a la HCE y al informe radiológico [30]: la citación de un paciente para adquirir una imagen (FT1, azul), el informado radiológico (FT2, verde) y el acceso del

especialista a la HCE e informes radiológicos (FT3, rojo). Por otra parte, la figura 3 muestra la secuencia ordenada de pasos de estos flujos de trabajo.

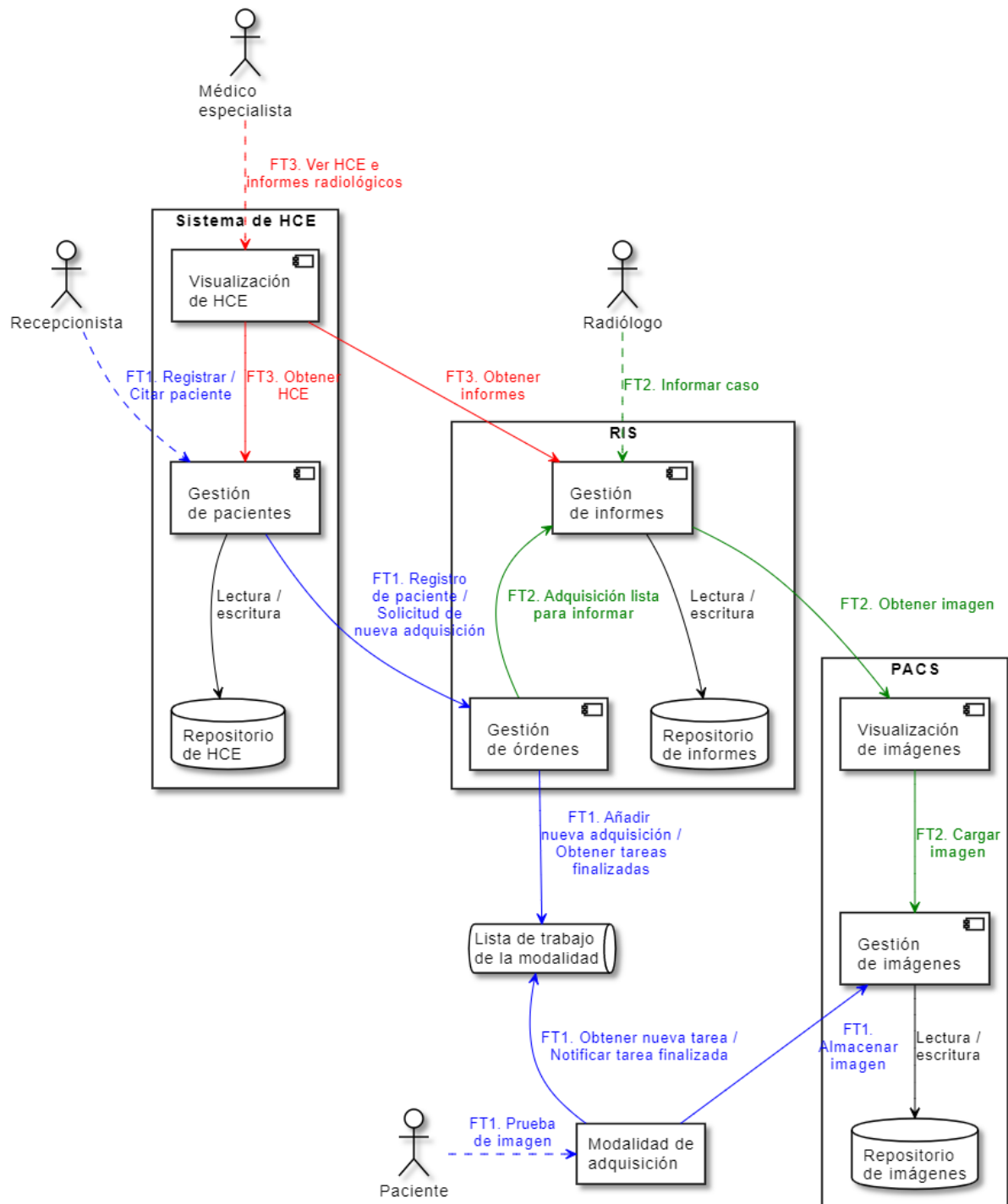


Figura 2. Diagrama de componentes de los sistemas de información sanitarios de HCE y radiología (HIS, RIS, PACS y modalidad de adquisición). Se describen tres flujos de trabajo (FT) que escenifican la interacción entre componentes: citación de paciente para adquirir una imagen (FT1, flechas azules), informado radiológico de la imagen (FT2, flechas verdes), Y acceso del especialista a la HCE e informes radiológicos del paciente (FT3, flechas rojas).

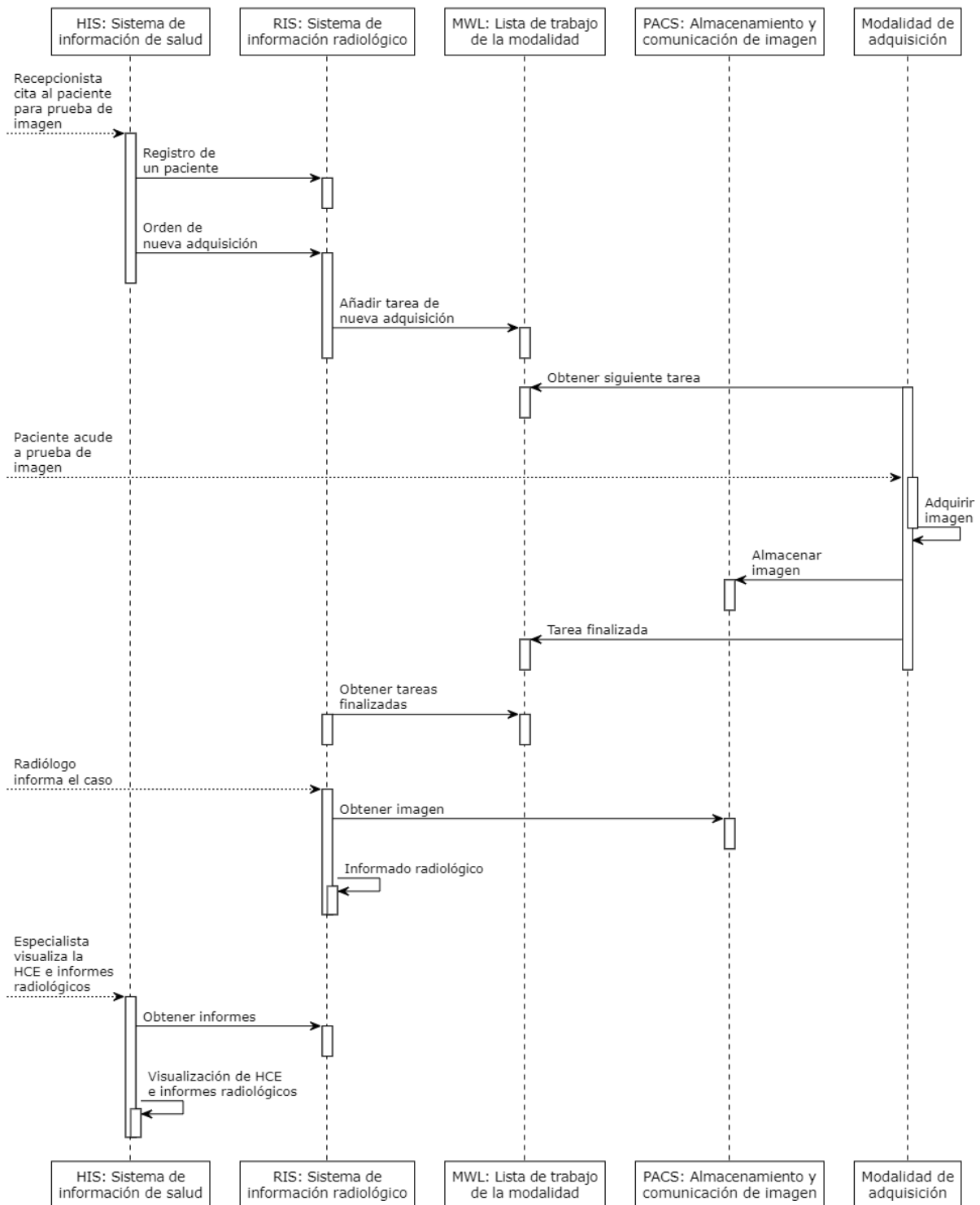


Figura 3. Diagrama de secuencia de los flujos de trabajo de citación para prueba de imagen médica, informado radiológico y visualización de HCE e informes radiológicos.

La integración de los sistemas descritos, dentro y fuera de los límites de un hospital, requiere del uso de formatos de datos, protocolos de comunicación y terminologías estándar que garanticen el correcto intercambio de datos entre las partes con un alto grado de interoperabilidad [31].

2.2. Historia clínica electrónica

La organización internacional de normalización ISO define la HCE como [32]: "Repositorio de información sobre el estado de salud de un sujeto de asistencia, en un formato computable, almacenado y transmitido de forma segura y accesible por múltiples usuarios autorizados. Su principal objetivo es proporcionar una asistencia sanitaria continua, de calidad y eficiente, que contenga información retrospectiva, actual y prospectiva.". La introducción de la norma ISO13606 ofrece otra definición [33]: "La HCE es el registro persistente, longitudinal y potencialmente multi-empresa o multinacional de la salud y la prestación de atención relativa a un único sujeto de la atención (el paciente), creado y almacenado en uno o más sistemas físicos con el fin de informar la futura atención de salud del sujeto y proporcionar un registro médico-legal de la atención que se ha prestado.". De acuerdo con estas definiciones, las HCE contienen una gran cantidad de datos biomédicos, con potencial para mejorar la atención sanitaria de las personas y proporcionar datos esenciales para mejorar la salud de la población. En cada visita clínica, los profesionales médicos capturan una instantánea de la salud de una persona. La sucesión de instantáneas de salud registradas en cada visita clínica constituye la HCE, el conjunto de información de salud de una persona a lo largo de su vida, desde el nacimiento hasta la muerte. Estos registros de salud se acumulan en repositorios masivos que incluyen datos estructurados (por ejemplo, variables numéricas o categóricas) y no estructurados (por ejemplo, extractos de texto libre en lenguaje natural). A medida que ha aumentado el tamaño, alcance y calidad de los datos de salud, han surgido cada vez más oportunidades para investigar cuestiones biológicas y médicas en nuevas modalidades y a una escala sin precedentes [34].

2.3. Estándares de historia clínica electrónica

En la década de los noventa se detectó la necesidad de una representación genérica de la HCE para su compartición entre sistemas. La gestión integrada de servicios sanitarios y continuidad en los cuidados médicos, requieren la adopción de normas y protocolos de comunicación, formatos, codificación y una estructura de HCE que permitan el intercambio de información sanitaria entre sistemas abiertos, distribuidos, interconectados e interoperables con un alto grado de fiabilidad y seguridad. Una HCE normalizada de acuerdo a estos criterios permite a los investigadores disponer de datos clínicos de más calidad y precisión, mitigando la redundancia de información y favoreciendo la validación de los datos clínicos. En Europa se han desarrollado una sucesión de proyectos de I+D financiados por la Comisión Europea y dos generaciones de normas del comité europeo de normalización (CEN) producidas por el comité técnico 251 (TC251) antes de la norma experimental ENV 13606, adoptada posteriormente por ISO (International Standard Organization). La intención de estos

proyectos y normas ha sido definir las características comunes de la HCE y adaptar los modelos de información y mensajes interpretables entre sistemas clínicos a las mismas. Es decir, permitir a diferentes sistemas clínicos intercambiar la HCE de un paciente, de forma normalizada y con capacidad para representar estos datos y su organización [35].

El objetivo de los estándares de HCE es desarrollar normas que permitan modelar, representar y comunicar datos de salud. Además, dado que la incertidumbre y el carácter incompleto son inherentes al ámbito de la salud, estos estándares pretenden encontrar el equilibrio entre ser lo suficientemente específicos para su ámbito actual, así como ser lo suficientemente genéricos y flexibles para dar cuenta de los conocimientos futuros. La HCE contiene un registro de la atención sanitaria prestada a una persona durante toda su vida. Su principal objetivo es facilitar y documentar la atención sanitaria distribuida en múltiples lugares. Aunque la investigación sobre HCE tiene mucho que ver con los datos y la gestión de los mismos, también tiene que ver, cada vez más, con el hecho de compartirlos. Es necesario tener la información bien definida y estructurada, y asimismo es necesario poder intercambiar información sanitaria sin problemas entre diferentes sistemas.

Las últimas generaciones de estándares para la representación y comunicación de HCEs se han diseñado para ser flexibles a la hora de representar las estructuras de datos clínicos y la información de contexto que acompaña a los mismos. Las estructuras de datos de HCE no son fijas. Pueden adaptarse para soportar nueva información y casos de uso. Para lograrlo, los nuevos estándares de la HCE definen modelos de datos genéricos, mínimamente acoplados a conceptos y estructuras de información específicas del ámbito sanitario. Solo representan las características comunes y mínimas de los datos clínicos. La mayor parte de la complejidad de los datos del ámbito sanitario está representada por modelos de información clínica complementarios. Por otra parte, contemplan una representación completa de la información de contexto. Es decir, cualquier metadato (dato acerca del dato) que informa sobre cómo se registraron los datos, o que influye en la interpretación de los datos clínicos contenidos en la HCE. Algunos de estos estándares se basan en el modelo dual de HCE, consistente en definir y proporcionar dos modelos bien diferenciados para la construcción y compartición de HCE. Por una parte, el modelo de referencia (MR) introduce propiedades estables y genéricas de la HCE, un marco común de componentes y especificaciones de cómo combinarlos para crear estructuras de datos más complejas, así como estructuras genéricas para albergar información contextual que debe acompañar a los datos para cumplir requisitos legales, éticos y de procedencia. La generalidad del MR se complementa con la particularidad del modelo de arquetipos (MA), que permite modelar arquetipos de conceptos de dominio clínico haciendo uso de los componentes del MR. Esto incluye definiciones formales de conceptos clínicos vinculados con recursos de conocimiento, y combinación jerarquizada de estructuras del MR restringidas para ajustarse a la definición

del concepto clínico modelado. Específicamente, los arquetipos restringen los valores de los atributos, estableciendo la opcionalidad, multiplicidad, tipo de datos y rango de valores en cualquier punto de la jerarquía. La descripción formal de los conceptos del arquetipo se consigue mediante el enlazado de las estructuras de datos y contenidos con terminologías y ontologías. La separación de modelos introducida por el modelo dual permite separar la implementación del sistema de HCE de la variabilidad de los conceptos clínicos. Así, cuando los conceptos clínicos cambian o evolucionan, el software no necesita ser modificado porque está diseñado para interpretar los componentes del MR, que permanecen inalterados cuando los arquetipos evolucionan para adaptarse a los cambios de los conceptos clínicos. Esta característica facilita la separación de responsabilidades, delegando a los técnicos el desarrollo software para dar soporte al MR, mientras los perfiles clínicos se centran en el modelado de conceptos clínicos haciendo uso del MA. Además, para facilitar la evolución de los arquetipos, las normas basadas en modelo dual de HCE contemplan mecanismos de reutilización de arquetipos basados en composición, especialización y versionado. La composición implica la combinación de diferentes arquetipos, incluyendo unos arquetipos en otros mediante referencias. La especialización consiste en restringir un arquetipo existente, refinando el concepto clínico que representa, pero manteniendo compatibilidad con el arquetipo inicial. Por último, el versionado introduce un mecanismo de evolución del concepto clínico representado, no necesariamente compatible con su versión previa.

A continuación, se describen algunos de los estándares y especificaciones técnicas de información sanitaria más relevantes [36]. El estándar ISO13606 se presenta con mayor detalle, al ser parte esencial del desarrollo de esta tesis.

2.3.1. ISO13606

El ámbito de aplicación de ISO13606 es comunicación e interoperabilidad semántica de los extractos de HCE [33]. En ISO13606 un extracto de HCE, unidad de comunicación, es una vista parcial o completa de la HCE de una persona y su información contextual, que se recoge para su comunicación entre diferentes sistemas de HCE. Esta norma define un MR que especifica las características básicas y la estructura de los datos clínicos (componentes del MR), y un MA, que permite modelar conceptos clínicos restringiendo los componentes del MR (nótese que también es posible referirse a los componentes del MR como clases abstractas). A continuación, se describen los componentes del MR de la norma ISO13606 (figura 4):

- EHR_EXTRACT: Clase abstracta de información clínica que representa la HCE de un paciente. Se trata del contenedor raíz de toda la HCE de un individuo. Incluye información sobre configuración de control de acceso, seguimiento del estado y versionado, un registro de los posibles cambios que se hayan dado en la HCE.

- FOLDER: Componente organizador jerárquico de alto nivel dentro del EHR_EXTRACT. Agrupa uno o más componentes COMPOSITION de acuerdo a diferentes criterios de ordenación.
 - Ejemplos: Obstetricia, atención de diabetes, pediatría, episodios del año anterior.
- COMPOSITION: Componente contenedor de información clínica de alto nivel subordinado a FOLDER. Se corresponde con documentos clínicos comúnmente utilizados, que incluyen la información emitida por un agente como resultado de un encuentro clínico. Puede contener componentes SECTION o ENTRY.
 - Ejemplos: Informe de alta, revisión semana 28 de embarazo.
- SECTION: Componente organizador jerárquico que proporciona una estructura lógica y de navegación para los datos dentro de un componente COMPOSITION. Refleja el flujo de recogida de información durante una sesión clínica de forma estructurada para facilitar su interpretación. Permite agrupar uno o más componentes ENTRY.
 - Ejemplos: Motivo de la sesión clínica, antecedentes, información de alergias.
- ENTRY: Clase abstracta para registrar declaraciones clínicas. Incluye toda la información registrada en la HCE como resultado de una observación clínica, evaluación (interpretación clínica), instrucción (intención), o acción.
 - Ejemplo: Un síntoma, una observación, un medicamento prescrito.
- CLUSTER: Clase organizadora concebida para representar estructuras de datos complejas como datos anidados, series temporales o tablas.
 - Ejemplos: Resultados de audiograma, interpretación de un electroencefalograma, diagnósticos diferenciales ponderados.
- ELEMENT: Componente de información clínica que representa un nodo hoja de la estructura de la HCE y contiene un único valor expresado mediante el tipo básico DATA_VALUE. Se entiende como tipo básico un dato único expresado mediante una cadena de caracteres, un número entero, o un número decimal, entre otros.
 - Ejemplos: Ritmo cardiaco, presión sanguínea sistólica, altura del paciente.

La figura 4 muestra los componentes del MR de la norma ISO13606 y las relaciones jerárquicas entre los mismos. A cada clase de información clínica le corresponde una clase organizadora. El componente de información clínica COMPOSITION cuenta con la clase organizadora SECTION para estructurar su contenido, mientras que ENTRY, clase de información clínica, cuenta con el componente organizador CLUSTER para ordenar sus datos clínicos. Al más alto nivel, para el componente de información clínica extracto de HCE (EHR_EXTRACT), se dispone de la clase organizativa FOLDER para organizar su información. La utilización de clases organizadoras es opcional, su función es únicamente la

de simplificar la interpretación de los datos. Opcionalmente, además, los componentes del MR organizadores pueden contener uno o más componentes de su mismo tipo. El Sistema Nacional de Salud español ha publicado un Conjunto Mínimo De Informes Clínicos (CMDIC) [37], modelado de acuerdo a la norma ISO13606, que configura la Historia Clínica Digital del Sistema Nacional de Salud (HCDSNS).

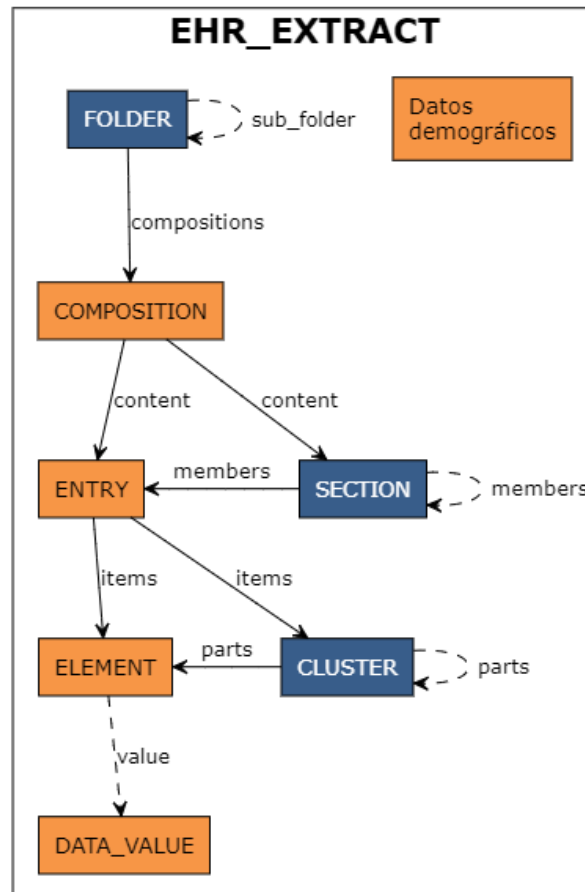


Figura 4. Componentes más relevantes del MR de la norma ISO13606. Distingue entre las clases orientadas a representar contenido clínico (EHR_EXTRACT, COMPOSITION, ENTRY, ELEMENT), en color naranja, y las clases que sirven para organizar o agrupar dicho contenido (FOLDER, SECTION, CLUSTER), en color azul.

2.3.2. OpenEHR

Especificación abierta para la creación de sistemas de HCE, incluyendo el almacenamiento, recuperación e intercambio de datos sanitarios, e información contextual básica necesaria para interpretar y reutilizar con seguridad los datos clínicos registrados [38]. OpenEHR define un MR y MA, similar al de la norma ISO13606, pero introduciendo algunas especializaciones de sus componentes del MR. Concretamente, la clase abstracta de información clínica ENTRY se puede especializar en algunas de las siguientes clases según su contenido:

- **OBSERVATION:** Información sobre un fenómeno observado, exámenes clínicos, pruebas de laboratorio y otros tipos de medidas objetivas.

- EVALUATION: Información de evaluaciones clínicas y diagnósticos.
- INSTRUCTION: Intervenciones concretas que deben ocurrir en el futuro, como por ejemplo planificación del cuidado u órdenes de medicación.
- ACTION: Información generada como resultado de la ejecución de acciones INSTRUCTION. Incluye declaraciones sobre acciones que ya se han realizado.
- ADMIN_ENTRY: Datos de índole administrativos.

2.3.3. HL7 CDA

HL7 (Health Level Seven) es un organismo internacional de estandarización, que entre otras cosas se ocupa de publicar normas para la representación, persistencia y comunicación de documentos clínicos como HL7 CDA [39]. Se trata de una representación, legible por el ser humano y computable por la máquina, de los datos clínicos generados durante los encuentros clínicos, o debido a una actividad de documentación. Los documentos HL7 CDA se basan en un Modelo de Información de Referencia (RIM) genérico, que incluye una cabecera con información de metadatos y un cuerpo con los datos clínicos. El cuerpo incluye una representación legible de los datos, pero también puede extenderse para estructurar los datos en secciones y declaraciones clínicas estructuradas (Acto, Observación, Administración de sustancias, Encuentro, Procedimiento, etc.).

2.3.4. HL7 FHIR

Marco estándar para la creación de una infraestructura de servicios web para la comunicación de información clínica [40]. HL7 FHIR se centra en la implementación, proporcionando la definición de porciones modulares de información clínica (recursos) para una rápida y fácil implementación utilizando estándares web, por ejemplo, utilizando servicios REST como infraestructura tecnológica para el intercambio de datos clínicos. HL7 ha apoyado predominantemente esta última faceta a través de la mensajería. En concreto, dispone del modelo unificado de acción y servicio USAM (Unified Service Action Model) que especifica las relaciones entre actos y entidades de dominio clínicos. Sin embargo, esta información se encuentra dentro de un marco de mensajería, y debe tener soporte para capturar el contexto y el estado de la progresión de la atención, así como los mensajes relevantes de forma persistente.

2.4. Estándares de imagen médica

El estándar para la comunicación y gestión de imágenes médicas DICOM [41] (Digital Imaging and Communications in Medicine), publicado y promovido por la asociación nacional de fabricantes electrónicos norteamericana NEMA, describe protocolos de red para la

integración de distintos dispositivos en el área de la imagen médica. Además, DICOM proporciona una especificación de formato adecuado para registrar, almacenar y compartir todos los tipos de imagen, señales médicas e informes radiológicos. Se trata del estándar más extendido para gestionar todo tipo de imágenes y señales médicas. Toda modalidad de adquisición de imagen médica contemporánea soporta el formato DICOM (por ejemplo, tomógrafos, ecógrafos, equipos de resonancia magnética o de rayos X). DICOM utiliza TCP como protocolo de transporte de red, y define una serie de operaciones para el almacenamiento, consulta, impresión de imágenes, y administración de listas de trabajo que permite programar tareas de los sistemas de información radiológicos y modalidades de adquisición, entre otros [42].

DICOM tiene cuatro niveles de organización que dan contexto a los archivos DICOM (figura 5), estos niveles son:

- Paciente: Sujeto al que pertenecen los estudios. El archivo DICOM contiene información demográfica del paciente como su nombre, sexo, identificador y fecha de nacimiento.
- Estudio: Especifica la información de un estudio realizado a un paciente. En general se corresponde a una única modalidad como tomografía axial computarizada (TAC) o imagen de resonancia magnética (IRM). Un estudio puede contener una o más series.
- Serie: Organiza objetos dentro de un estudio. Por ejemplo, dentro de una tomografía computarizada que incluye varios cortes, una serie podría corresponder a las imágenes realizadas con cortes en cierto ángulo y otra serie a los cortes realizados en otro ángulo. En el caso de una radiografía simple, la serie se compone de una única imagen.
- Objeto: Puede representar una imagen, video, señal, sonido o informe. Este último es el resultado de la evaluación de la imagen por parte del radiólogo especialista. Existe un quinto nivel de organización, el marco o frame, donde un objeto puede contener múltiples frames, por ejemplo en el caso de videos, cada frame corresponde a un fotograma. A los objetos con múltiples marcos se les llama multi-frame.

Los archivos DICOM están divididos en una cabecera y un cuerpo. La cabecera está estructurada en un formato de clave-valor, donde las claves son conocidas como etiquetas DICOM y para cada etiqueta se corresponde un valor. Cada etiqueta tiene un número identificador y un nombre. En la especificación de DICOM se define el significado de las etiquetas públicas, aunque también se reserva un rango de etiquetas privadas para los fabricantes. El cuerpo puede incluir uno o más bloques de datos (imágenes, videos, etc.). En la tabla 2 se muestra el identificador, nombre y descripción de algunas de las etiquetas DICOM públicas más utilizadas.

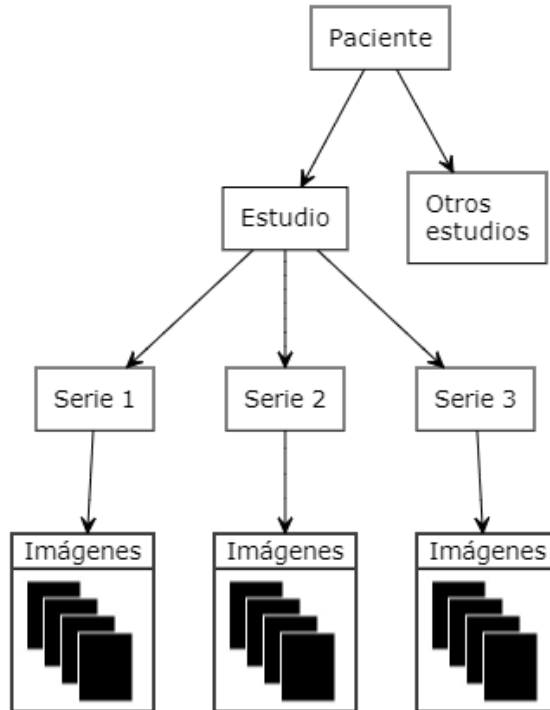


Figura 5. Jerarquía de niveles de información de los objetos DICOM. Cada objeto DICOM incluye información sobre la imagen, serie, estudio y paciente al que pertenece. Dicha información permite agrupar las imágenes para visualizar cada secuencia adquirida.

Etiqueta	Nombre	Descripción
0008,0020	StudyDate	Fecha de realización del estudio
0008,0050	AccessionNumber	Identificador de la orden de adquisición del estudio
0008,0060	Modality	Modalidad utilizada para realizar el estudio
0008,0080	InstitutionName	Institución donde se realizó el estudio
0010,0010	PatientName	Nombre y apellido del paciente
0010,0020	PatientID	Identificador del paciente
0010,0030	PatientBirthDate	Fecha de nacimiento del paciente
0010,0040	PatientSex	Sexo del paciente
0020,000D	StudyInstanceUID	Identificador del estudio
0020,000E	SeriesInstanceUID	Identificador de la serie
0008,0018	SOPInstanceUID	Identificador del objeto

Tabla 2. Identificador, nombre y descripción de etiquetas públicas de la cabecera DICOM ampliamente utilizadas.

DICOM se emplea para comunicar información de estudios de imagen dentro del departamento de radiología, entre el RIS, el PACS y las modalidades de adquisición. DICOM plantea el escenario de sistemas de información en radiología como una red de nodos, donde cada uno de ellos puede actuar como servidor o cliente según la operación a realizar. Por ejemplo, un RIS actúa como cliente cuando consulta al PACS acerca de los estudios de un determinado paciente, y al mismo tiempo actúa como servidor cuando es consultado por los informes radiológicos de un paciente. Los servicios u operaciones de comunicación DICOM son muy variados, siendo los más destacables el de enviar, consultar y recuperar imágenes. La especificación de la norma determina siempre un par de componentes para cada uno de estos servicios: la clase proveedor y la clase cliente. Así, cada sistema de información que deba involucrarse en dichas operaciones deberá implementar, al menos, una de estas dos clases.

2.5. SADC basados en guías clínicas y biomarcadores de imagen

Los biomarcadores de imagen son parámetros cuantitativos anatómicos, fisiológicos, bioquímicos o moleculares, que pueden obtenerse mediante SADC basados en procesamiento de imágenes y señales médicas, para establecer la presencia o gravedad de las enfermedades. Los biomarcadores de imagen desempeñan un papel cada vez más importante en la investigación médica y la práctica clínica en la era de la medicina de precisión [43]. Existe una alta heterogeneidad de tipos de biomarcadores de imagen en función del tipo de imagen para el que estén diseñados (por ejemplo, resonancia o tomografía), la patología o característica fisiológica que cuantifican (por ejemplo, permeabilidad de fluidos entre tejidos, detección de lesiones, alteraciones morfológicas de órganos), el uso para el que han sido diseñados (por ejemplo, cálculos semi-automáticos para la estadificación de lesiones durante el informado radiológico o sistemas sin supervisión de vigilancia de posibles casos patológicos en cribados de urgencias), o atendiendo al tipo de ayuda al diagnóstico proporcionada y la metodología empleada para su obtención. Los biomarcadores de imagen pueden cuantificar una sustancia rastreable introducida en un organismo como medio para examinar la función de un órgano, un producto bioquímico derivado de los fluidos corporales (a menudo utilizado en ensayos clínicos), una señal anatómica o una imagen médica de una parte del cuerpo. Su uso es común en la práctica clínica, en el contexto de la medicina personalizada, para describir condiciones normales y/o patológicas, donde el uso de biomarcadores de imagen puede ayudar a identificar a los pacientes que tienen más probabilidades de responder favorablemente a una determinada terapia [12]. Su poder

predictivo o su capacidad en métodos de imagen cuantitativos son solo algunas características que demuestran su potencial.

Los biomarcadores de imagen proporcionan ayuda a la decisión clínica y ofrecen al radiólogo información objetiva derivada de la adquisición del paciente, que contribuye a homogeneizar criterios de diagnóstico radiológico. De la misma manera, las guías clínicas de radiología establecen pautas comunes (por ejemplo, criterios de puntuación de la gravedad de las lesiones en función de sus características) para la evaluación de determinadas patologías, como las lesiones en próstata, mama o hígado, entre otras áreas anatómicas. El primer prototipo de estas guías, el sistema de datos e informado de imágenes de mama BI-RADS, fue publicado por el colegio americano de radiología ACR en 1993 para abordar la falta de uniformidad en los informes de mamografía. Actualmente, hay nueve sistemas con el pleno respaldo del ACR y varios sistemas nuevos en diversas fases de desarrollo [44] (tabla 3):

- BI-RADS: Cáncer de mama, diagnosticado mediante mamografía, IRM o ultrasonidos (US).
- C-RADS: Cáncer de colon, diagnosticado con colonoscopia virtual basada en TAC.
- LI-RADS: Cáncer de hígado; diagnosticado mediante IRM, TAC, US o US con contraste.
- Lung-RADS: Cáncer de pulmón, diagnosticado con TAC de baja radiación.
- NI-RADS: Cánceres de cabeza y cuello, diagnosticados mediante tomografía por emisión de positrones (PET/TAC) o IRM.
- O-RADS: Masas anexiales ováricas, diagnosticadas mediante US.
- PI-RADS: Cáncer de próstata, diagnosticado con IRM.
- TI-RADS: Cáncer de tiroides, diagnosticado mediante US.
- CAD-RADS: Enfermedades arteriales coronarias, diagnosticadas con angiografía / TAC.

La tabla 3 recoge un resumen de las guías clínicas de radiología más utilizadas y sus principales características: condición patológica, rango de puntuaciones o categorías, criterio base para la asignación de puntuaciones y modalidades de imagen sobre las que se aplica.

Guía clínica radiológica	Condición patológica	Puntuación o categorización	Criterio de asignación de puntuación	Modalidad de imagen
BI-RADS	Cáncer de pecho	0-6 (4A-C)	Basado en el paciente	Mamografía, IRM, US
C-RADS	Cáncer de colon	Hallazgos en el colon (C0-4); hallazgos extracolónicos (E0-4)	Basado en la lesión	Colonoscopia virtual (TAC)
CAD-RADS	Enfermedad de arterias coronarias	0-5, N	Basado en el paciente	Angiografía por TAC
LI-RADS	Cáncer de hígado (hepatocarcinoma)	1-5, TIV, NC, M, TR	Basado en la lesión o basado en el paciente	TAC, IRM, US con contraste, US
Lung-RADS	Cáncer de pulmón	0-4 (4A, B, X), S, C	Basado en el paciente	TAC de baja radiación
NI-RADS	Cánceres de cabeza y cuello	0-4 (2A, B)	Basado en el paciente	PET-TAC, TAC, IRM
O-RADS	Masas anexas al ovario	N/A	N/A	US
PI-RADS	Cáncer de próstata	1-5	Basado en la lesión	IRM multi-paramétrica
TI-RADS	Cáncer de tiroides	1-5	Basado en la lesión	US

Tabla 3. Resumen de nueve sistemas de datos e informado RADS (Reporting And Data System) publicados por el ACR. C = cáncer de pulmón previo; M = maligno pero no hepatocarcinoma; N, NC = no diagnóstico; S = hallazgos clínicamente significativos o potencialmente significativos (no cáncer de pulmón); TIV = tumor en vena; TR = tratado.

A continuación, se introducen los tipos de SADC basados en biomarcadores de imagen y guías clínicas de radiología más relevantes en el contexto de esta tesis, citando algunos ejemplos que ayudan a entender el trabajo de investigación presentado en los siguientes capítulos.

2.5.1. SADC basados en reglas

Los SADC basados en reglas son la herramienta adecuada para modelar sistemas de puntuación de guías clínicas de radiología, y asistir al radiólogo en el proceso de evaluación del caso. Por ejemplo, el sistema de puntuación de la guía clínica de radiología PI-RADS define una escala de 5 puntos para indicar la probabilidad de presencia de un cáncer de próstata clínicamente significativo. Para ello, el radiólogo debe evaluar diferentes secuencias de IRM y valorarlas de acuerdo a criterios cualitativos específicos. Dichas secuencias son: imagen T2 ponderada o T2W (T2 Weighted), imagen de difusión ponderada o DWI (Diffusion Weighted Image) e imagen de perfusión con contraste dinámico mejorado o DCE (Dynamic Contrast Enhanced). La DWI es la secuencia dominante en la evaluación de la zona periférica (ZP). Esto significa que la puntuación de PI-RADS para la ZP se basa en la DWI, a menos que la puntuación de DWI sea PI-RADS 3. En este caso se utiliza la DCE para decidir entre PI-RADS 3 (DCE negativa - sin realce claro) y PI-RADS 4 (DCE positiva - se observa realce focal o temprano de contraste). La guía PI-RADS continúa estableciendo criterios similares para las puntuaciones de T2W, DWI y DCE, que cambian según la zona de la próstata donde se identifique la lesión. La evaluación PI-RADS no es una suma de las puntuaciones de T2W, DWI y DCE, sino una correlación de las puntuaciones cualitativas de las diferentes secuencias de IRM de próstata [45]. Los SADC basados en reglas son adecuados para modelar este comportamiento, facilitando el cálculo de la puntuación PI-RADS final basándose en las puntuaciones cualitativas individuales de T2W, DWI y DCE para los diferentes hallazgos susceptibles de ser potenciales lesiones prostáticas. Finalmente, las puntuaciones PI-RADS 1 y PI-RADS 2 indican que el cáncer de próstata clínicamente significativo es altamente improbable o poco probable, mientras que PI-RADS 4 y PI-RADS 5 muestran una posibilidad alta o muy alta de cáncer clínicamente significativo [46]. De igual manera, los SADC basados en reglas son adecuados para el modelado de otros sistemas de puntuación basados en guías clínicas de radiología, como los que se listan en la tabla 3.

2.5.2. SADC basados en visión por computación

El procesado de imágenes por computación consiste en transformar una imagen en otra imagen, habitualmente con el objetivo de eliminar o realzar algunas características. El análisis de imagen, además, extrae parámetros de la imagen para su evaluación. En medicina, la

visión por computación combina el procesado cuantitativo y simbólico de la imagen para proporcionar ayuda al diagnóstico clínico, cuantificando variables de potencial interés por ser indicadores directos o indirectos de posibles condiciones patológicas. Las motivaciones para utilizar visión por computación en el ámbito de la imagen médica abarcan el análisis de nuevas modalidades y multimodal, análisis morfológico de estructuras anatómicas, la mejora de la interpretación clínica de la imagen mediante parámetros cuantitativos objetivos, el aumento de la precisión en el diagnóstico, la automatización de procesos o la obtención de conclusiones a partir de datos volumétricos [47]. Un ejemplo de SADC basados en visión por computación son los parámetros cuantitativos empleados para la evaluación de IRM de próstata, entre los que se incluyen el coeficiente de difusión aparente ADC (Apparent Diffusion Coefficient), y los modelos farmacocinéticos que permiten cuantificar distintos parámetros relacionados con la angiogénesis. El ADC se calcula sobre secuencias DWI con múltiples b-value¹, dando lugar a la imagen del mapa de ADC mediante el mapeo de píxeles. Este sirve para incrementar la sensibilidad y especificidad en más de un 54% en la detección del cáncer de próstata por IRM, cuando se utiliza en conjunción con la secuencia T2W [48]. Por otro lado, los modelos farmacocinéticos se focalizan en parametrizar la permeabilidad y el flujo sanguíneo, cuantificando secuencias de DCE [49]. Los más relevantes son K^{trans} , constante de permeabilidad capilar, que cuantifica el paso de contraste a través del endotelio desde el compartimento vascular al intersticio. k_{ep} , Coeficiente de extracción, modela la vuelta del contraste extravasado al espacio vascular. Y v_e , fracción de volumen del espacio extracelular ocupada por el tumor [50]. Todos estos parámetros se pueden representar gráficamente mediante mapas paramétricos, siendo especialmente relevantes en determinadas regiones de interés dentro de las IRM de próstata. Para delimitar la región de la próstata y sus secciones, se emplean técnicas de procesado de imágenes sobre la secuencia T2W, donde las estructuras anatómicas se distinguen con más claridad [51]. En su conjunto, los parámetros descritos junto con el proceso de segmentación de la próstata constituyen un complejo sistema de ayuda a la decisión clínica para el diagnóstico de lesiones de próstata potencialmente tumorales. Procedimientos similares se emplean en otras áreas anatómicas y modalidades de adquisición, para construir SADC basados en biomarcadores de imagen con técnicas de visión por computación, que asisten el proceso de diagnóstico de diversas patologías [52].

¹ b-value: Factor que refleja la fuerza y el tiempo de los gradientes utilizados para generar imágenes ponderadas por difusión. Cuanto mayor sea el valor b, más fuertes serán los efectos de la difusión.

2.5.3. SADC basados en aprendizaje automático

En el contexto del procesamiento de imagen médica, las técnicas de aprendizaje automático permiten abordar desafíos que no son viables mediante las metodologías clásicas deterministas de visión por computador. Las redes neuronales convolucionales (Convolutional Neural Networks, CNN) están especialmente adaptadas para el manejo de datos bidimensionales y han tenido también un gran impacto en otros dominios, como es el caso del reconocimiento de voz. Las arquitecturas de estas CNN pueden incluir un elevado número de capas ocultas entre la entrada y la salida, que hacen que a esta tecnología se la conozca con el nombre de aprendizaje profundo o deep learning. La detección de lesiones, la segmentación automática de órganos y la clasificación de imágenes en patologías y sus subgrupos, son algunas de sus funcionalidades más destacables en el ámbito de la imagen médica [53]. Las CNN aprenden a identificar aquellas estructuras susceptibles de corresponder con una lesión, obviando estructuras anatómicas normales con intensidades similares. De igual manera, órganos y tejidos como la próstata, el hígado o el cartílago de la rodilla en IRM pueden ser delimitados en pocos segundos mediante CNN, en contraposición con las técnicas de visión por computación, con las que resulta sumamente complejo segmentar la región de interés en todo el órgano. Las CNN aprenden de cómo los expertos delimitan una parte de la anatomía en un conjunto de tamaño considerable de datos etiquetados, siendo capaz de extraer la anatomía de cualquier nueva imagen que se le proporcione. En lo que respecta a los problemas de clasificación, el aprendizaje automático tiene el potencial de procesar una imagen radiológica y clasificarla en aquellas clases con las que la imagen presenta una mayor similitud, permitiendo distinguir imágenes de estructuras anatómicas patológicas de las sanas.

Técnicas de aprendizaje automático como las CNN se nutren de grandes volúmenes de datos y propiedades extraídos, entre otros medios, mediante técnicas de visión por computación. Dichos datos se conocen como radiómica. Los métodos de aprendizaje automático, permiten fusionar las propiedades radiómicas en un único valor, como por ejemplo una firma tumoral, que podría estar relacionada con la probabilidad del estadio de la enfermedad [54]. Así, los SADC basados en aprendizaje automático complementan a los SADC basados en visión por computación, dando lugar a sofisticadas soluciones avanzadas para el soporte a la decisión clínica.

2.6. Informe estructurado radiológico

Tradicionalmente, los informes radiológicos se elaboraban con un lenguaje narrativo de texto libre. Los estudios demuestran que el uso de informes con lenguaje narrativo puede suponer un obstáculo para la atención óptima del paciente. La excesiva variabilidad en el lenguaje, la

longitud y el estilo puede reducir al mínimo la claridad del informe, dificultando que los médicos destinatarios identifiquen la información clave necesaria para la atención del paciente. La presentación de informes estructurados como posible solución para mejorar la calidad de los informes radiológicos, y su clasificación en niveles básico, medio y alto ha sido sujeto de estudio en la literatura [55]. En su nivel básico, un informe estructurado debe organizarse con títulos, como historia clínica, indicación, técnica, hallazgos e impresión. En el siguiente nivel se organiza la sección de hallazgos con subtítulos, como los distintos órganos (o estructuras anatómicas) visualizados. En el nivel más alto, el informe tiene todas las características mencionadas anteriormente y utiliza un lenguaje estandarizado basado en un léxico ampliamente aceptado. Cada vez más, los centros académicos utilizan informes radiológicos estructurados que contienen plantillas, macros o listas de comprobación precargadas.

A los radiólogos en formación se les enseña a mirar las imágenes de forma sistemática y a describir sus hallazgos tal y como los ven. Al final de este proceso, formulan un diagnóstico que resume los hallazgos patológicos significativos y los correlacionan con la información y la pregunta clínica proporcionada por el médico remitente. Los radiólogos más experimentados elaboran informes más breves. La narrativa de los informes radiológicos, sin embargo, se mantiene principalmente porque han sido entrenados para ello. No cabe duda de que se puede cuestionar si la forma de mirar y razonar de un radiólogo tiene que ser el factor decisivo a la hora de presentar los resultados de un procedimiento de imagen al médico destinatario. El médico destinatario puede preferir un informe bien estructurado, un instrumento útil y reproducible para el diagnóstico y el seguimiento, en lugar de una aproximación particular del radiólogo. Además, los informes narrativos no siempre abordan las preguntas clínicas clave, pueden contener errores clínicamente importantes y términos ambiguos [56]. Determinadas áreas de especialidad se prestan con bastante facilidad al informado estructurado, como por ejemplo la mamografía de cribado, la IRM funcional, los estudios ortopédicos y los estudios de seguimiento de pacientes oncológicos que utilizan los criterios RECIST². Las plantillas de informes específicas para cada enfermedad pueden mejorar la claridad y calidad de los informes, y garantizar en la práctica un uso coherente de las terminologías clínicas y una documentación completa del informe radiológico, con el consecuente impacto positivo en la investigación radiológica al facilitar la extracción de datos de manera estructurada [57].

Un informe estructurado supone: la presencia de listas y relaciones jerárquicas; el uso de contenido codificado o numérico además de texto plano; el uso de relaciones entre conceptos; y la presencia de referencias incrustadas a imágenes y objetos similares [58]. La elaboración

² Método estándar que se usa para medir el modo en que un paciente de cáncer responde al tratamiento. Permite determinar si los tumores se reducen, permanecen igual o se agrandan.

de informes estructurados según estos criterios abre posibilidades inalcanzables con los informes de texto tradicionales. En el contexto de la asistencia sanitaria centrada en el paciente, es necesario que los informes puedan ser leídos por profesionales sanitarios de diferentes especialidades e incluso por profesionales no sanitarios (el paciente y sus cuidadores) [59]. El informe estructurado puede utilizarse para generar contenidos en lenguaje natural de acuerdo con el nivel de comprensión del lector al que va dirigido. Por otra parte, el razonamiento automatizado puede proporcionar a los radiólogos sugerencias de diagnóstico para casos inusuales, promoviendo la medicina basada en la evidencia al integrar los resultados de las herramientas de apoyo a la toma de decisiones en los informes radiológicos [60]. Los informes basados en listas de comprobación pueden reducir los errores de diagnóstico, así como los errores gramaticales de los sistemas digitales de reconocimiento de voz [61]. De hecho, si consideramos un informe estructurado como una guía o asistente a la hora de informar, podría ser útil de dos maneras: para los radiólogos experimentados, como ayuda para responder adecuadamente a la pregunta clínica, y para los radiólogos inexpertos, como solucionador de problemas en casos difíciles. Diversos estudios demuestran que es posible implantar el informe estructurado en el departamento de radiología con una amplia adopción [62], destacando también la preferencia de los radiólogos por los informes estructurados en contraposición al informe clásico [63].

Las soluciones actuales ponen de manifiesto las ventajas descritas con respecto al dictado convencional. Sistemas de plantillas de informe estructurado [64], integraciones con sistemas de puntuación cualitativa radiológicos [65], con resultados cuantitativos de biomarcadores de imagen [66], o con resultados de SADC para potenciar la medicina personalizada centrada en el paciente [67]. La explotación de estas soluciones requiere de una integración completa del informe estructurado con sistemas de información de HCE y radiología, utilizando estándares de HCE, el estándar DICOM y el perfil de integración para la empresa sanitaria IHE (Integrating the Healthcare Enterprise) de gestión de plantillas de informe radiológico MRRT (Management of Radiology Reporting Templates) [68]. Concretamente, el estándar DICOM especifica el formato DICOM-SR (DICOM Structured Report) y un protocolo de comunicación para los IRE, mientras que el perfil de integración IHE MRRT análogamente proporciona especificaciones de formato y protocolo de comunicación para plantillas de IRE. Es necesaria la codificación de los IRE en DICOM-SR, la compatibilidad de plantillas de IRE con el estándar IHE MRRT [69], y la implementación de plataformas de informado que cumplan con ambas normas [70], para implementar flujos de trabajo prácticos que hagan los datos de la rutina clínica más accesibles [71]. DICOM-SR define un MR para definir informes estructurados (figura 6), que permite modelar jerarquías estructuradas de elementos de contenido codificados con terminologías [72].

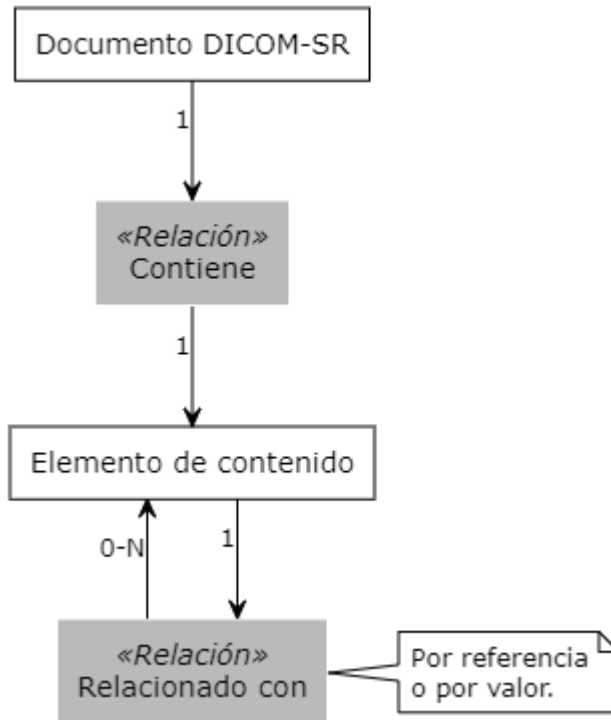


Figura 6. Modelo entidad-relación de DICOM SR. El modelo describe cómo el informe estructurado contiene un único elemento de contenido raíz, que a su vez puede estar relacionado con otros elementos de contenido.

Un elemento de contenido³ contiene o referencia a otros elementos de contenido. Las relaciones entre elementos de contenido pueden darse por valor (un elemento contiene a otro) o por referencia (un elemento hace referencia a otro que no desciende directamente de él). Opcionalmente, cada plantilla de IRE puede permitir o restringir las relaciones cíclicas por referencia. La figura 7 muestra el aspecto de una plantilla de IRE arbitraria, modelada de acuerdo al MR de DICOM-SR.

El estándar DICOM publica una serie de plantillas estándares para modelar el IRE en formato DICOM-SR [73]. Estas plantillas están basadas en el modelo de referencia descrito en la figura 6, y consisten mayoritariamente en estructuras de datos reutilizables para la construcción del IRE en DICOM-SR. Algunos ejemplos de dichas estructuras son: medidas numéricas, propiedades de medidas estadísticas, imagen o coordenadas espaciales, forma de onda o coordenadas temporales, contexto del procedimiento, contexto del sujeto, e incluso plantillas para el informe básico de diagnóstico por imagen y el informe en formato narrativo. El informe estructurado ha evolucionado desde plantillas comunes, definidas por comités de expertos como el panel consultor de librerías de plantillas TLAP (Templates Library Advisory Panel) [74], hasta sofisticados formularios procesables que incluyen validaciones,

³ Módulo de contenido de un documento de informe estructurado definido en el estándar DICOM (figura 2, adaptada de la norma DICOM [72]). Incluye información sobre el tipo de dato que contiene (TEXT, NUM, CODE, DATE, TIME, DATETIME, UIDREF, PNAME, COMPOSITE, IMAGE, WAVEFORM, SCOOD, SCOOD3D, TCOORD, CONTAINER), su valor y enlaces terminológicos.

calculadoras y enlaces a imágenes clave [75]. Se espera que en un futuro próximo se incorpore al informe estructurado información adicional procedente de Big Data, Inteligencia Artificial (IA) o SADC [76]. Los procedimientos para integrar biomarcadores de imagen en el informe estructurado acercan el paradigma de la medicina personalizada al flujo de trabajo radiológico. Los resultados de la cuantificación pueden complementar el diagnóstico radiológico, mejorando la precisión en el diagnóstico y la evaluación de la eficacia de los tratamientos [77].

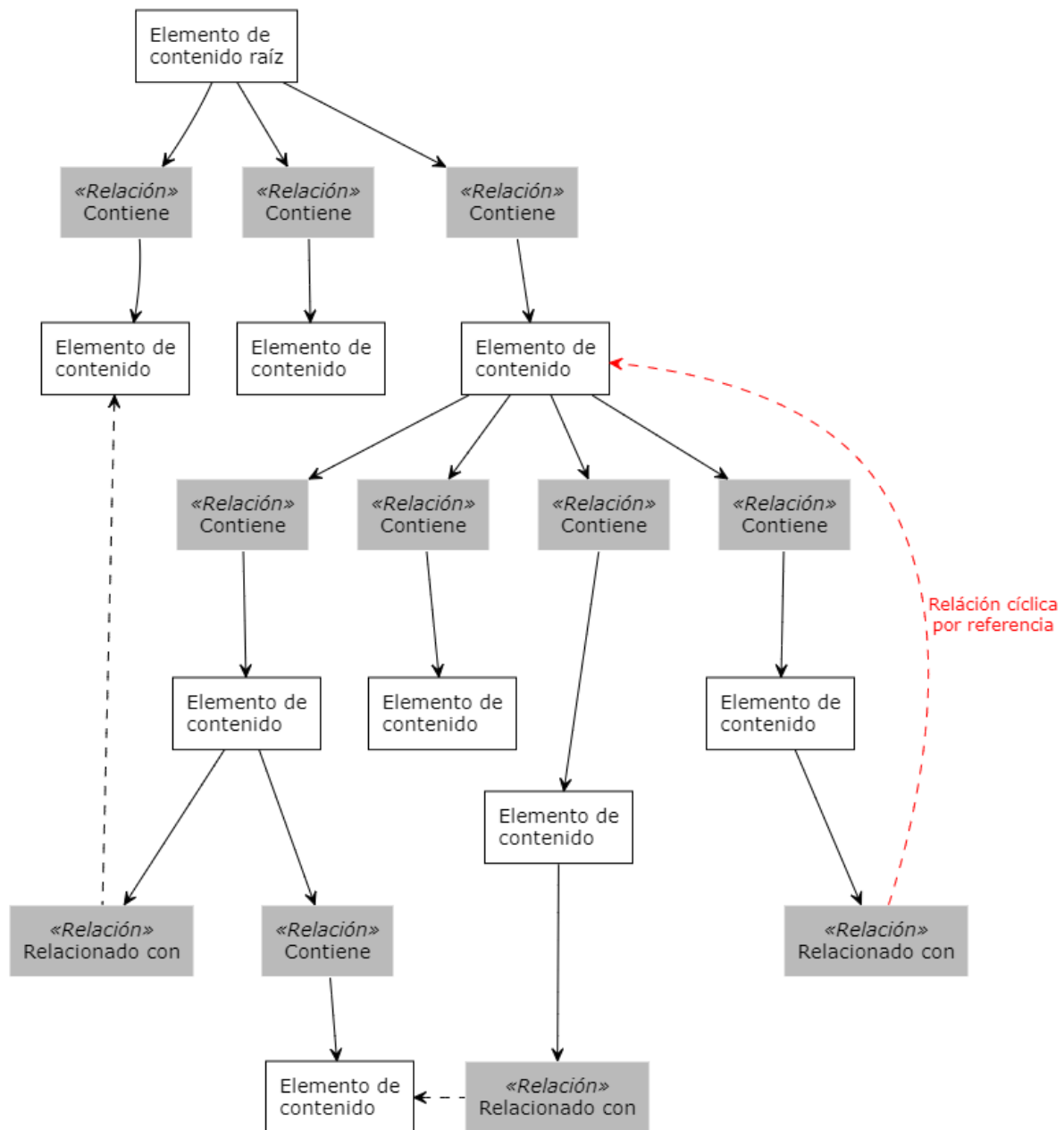


Figura 7. Ejemplo de plantilla de IRE de acuerdo al modelo de referencia DICOM-SR. Las relaciones entre elementos de contenido pueden ser por valor (flechas con línea continua) o por referencia (flechas con línea punteada). Opcionalmente, se pueden permitir o restringir las relaciones cíclicas por referencia.

A pesar de los beneficios descritos, existen limitaciones que deben ser superadas para la adopción generalizada de los informes estructurados [55]. (I) Los radiólogos pueden ser reacios a cambiar la forma convencional de informar, ya que la curva de aprendizaje del nuevo estilo de informe puede afectar negativamente al flujo de trabajo y a su productividad. (II) Las tasas de error podrían aumentar si los informes estructurados se utilizan de forma inadecuada (por ejemplo, si los valores por defecto de ciertos campos de la plantilla no han sido confirmados por el usuario). (III) La interrupción del patrón de búsqueda visual, al mantener los ojos centrados en el informe en lugar de en las imágenes, puede aumentar el tiempo de elaboración de informes. (IV) Las plantillas que incluyen información innecesaria o irrelevante pueden socavar la coherencia del informe y la interpretación del mismo por parte de los médicos remitentes. Por último, (V) la falta de soporte para múltiples idiomas puede dificultar la adaptación del informe al público objetivo.

2.7. Interoperabilidad de sistemas de información sanitarios

La sociedad de sistemas de información y gestión sanitaria norteamericana HIMSS (Healthcare Information and Management Systems Society) [78], define la interoperabilidad en salud [79] como la capacidad de diferentes sistemas de información y aplicaciones de software para comunicarse, intercambiar datos y utilizar la información que se ha intercambiado. Los esquemas y estándares de intercambio de datos deben permitir que los datos sean compartidos por los médicos, el laboratorio, el hospital, la farmacia y el paciente, independientemente del sistema de información o el proveedor del mismo. La interoperabilidad se refiere a la capacidad de los sistemas de información sanitaria para trabajar juntos dentro y fuera de los límites de la organización, con el fin de mejorar la salud de los individuos y las comunidades y hacer eficaz la prestación de la asistencia sanitaria.

La HIMSS define tres niveles de interoperabilidad de la tecnología de la información sanitaria:

1) Fundamental; 2) Estructural; y 3) Semántica.

1. La interoperabilidad fundamental permite que los datos de un sistema informático sean recibidos por otro y no requiere que el sistema informático receptor pueda interpretarlos.
2. La interoperabilidad estructural es un nivel intermedio que define la estructura o el formato del intercambio de datos (es decir, las normas de formato de los mensajes), en el que se produce un movimiento uniforme de los datos sanitarios de un sistema a otro, de manera que la finalidad y el significado clínico u operativo de los datos se mantienen inalterados. La interoperabilidad estructural define la sintaxis del intercambio de datos. Garantiza que los intercambios de datos entre sistemas informáticos puedan interpretarse a nivel de campo de dato.

3. La interoperabilidad semántica proporciona la interoperabilidad al más alto nivel, que es la capacidad de dos o más sistemas o elementos para intercambiar información y utilizar la información que se ha intercambiado. La interoperabilidad semántica utiliza, además de la estructura de los datos, su codificación. Mediante términos codificados, incluye el vocabulario para que los sistemas informáticos receptores puedan interpretar los datos. Este nivel de interoperabilidad admite el intercambio electrónico de datos financieros relacionados con la salud, datos de HCE creados por el paciente e información de resumen del paciente entre los cuidadores y otras partes autorizadas. La interoperabilidad semántica se construye a través de sistemas de HCE potencialmente dispares, sistemas de información sanitarios, dispositivos médicos, tecnologías móviles y otros sistemas para mejorar el bienestar, así como la calidad, la seguridad, la relación coste-efectividad y el acceso a la prestación de asistencia sanitaria.

La asociación EN13606 define la interoperabilidad como una propiedad, relacionada con la capacidad de diversos sistemas y organizaciones para trabajar juntos (interoperar). Además, la asociación hace referencia a la definición del IEEE, que introduce la interoperabilidad como la capacidad de dos o más sistemas o componentes para intercambiar información y utilizar la información que se ha intercambiado. Distinguen dos tipos de interoperabilidad: 1) Interoperabilidad sintáctica y 2) Interoperabilidad semántica.

1. Interoperabilidad sintáctica: Si dos o más sistemas son capaces de comunicarse e intercambiar datos, están mostrando interoperabilidad sintáctica. Los formatos de datos especificados y los protocolos de comunicación son fundamentales. En general, los formatos XML o JSON proporcionan interoperabilidad sintáctica. La interoperabilidad sintáctica es un prerrequisito para cualquier intento de interoperabilidad posterior.
2. Interoperabilidad semántica: Más allá de la capacidad de dos o más sistemas informáticos para intercambiar información, la interoperabilidad semántica es la capacidad de interpretar automáticamente la información intercambiada de forma significativa y precisa, con el fin de producir resultados útiles definidos por los usuarios finales de ambos sistemas. Para lograr la interoperabilidad semántica, ambas partes deben remitirse a un MR de intercambio de información común. Los datos en los intercambios de información están definidos de forma inequívoca: la interpretación de los datos es la misma en el emisor y el receptor.

El proyecto europeo SemanticHealthNet [80] desarrolla una hoja de ruta a largo plazo para la investigación y aplicación de soluciones efectivas de interoperabilidad semántica en salud,

establece los modelos de información, de arquetipos y de dominio (terminologías y ontologías médicas) como los 3 pilares básicos de la interoperabilidad semántica, y define los siguientes niveles de interoperabilidad entre SIS:

1. Interoperabilidad técnica y sintáctica: La institución es capaz de recibir documentos electrónicos sin capacidad de interpretación semántica. Las aplicaciones ampliamente disponibles que soportan la interoperabilidad sintáctica (como los navegadores web y los clientes de correo electrónico), permiten la descarga de los datos de los pacientes y proporcionan un acceso inmediato. Solo la intervención humana permite interpretar la información recibida para añadirla al sistema de información del hospital.
2. Interoperabilidad semántica parcial: Las secciones más relevantes de la HCE (como datos demográficos, alergias, diagnósticos y otras partes de la HCE) se codifican utilizando esquemas de codificación internacionales, que el sistema de información del hospital puede detectar automáticamente, interpretar y presentar de forma significativa al médico que utiliza los datos. Se distinguen dos niveles ortogonales de interoperabilidad semántica parcial:
 - a. Interoperabilidad semántica unidireccional
 - b. Interoperabilidad semántica bidireccional de secciones significativas.
3. Interoperabilidad semántica total, cooperatividad: En esta situación ideal, los sistemas son capaces de acceder, interpretar y presentar automáticamente toda la información médica necesaria al médico en el punto de atención. Ni las diferencias lingüísticas ni las tecnológicas impiden que el sistema integre sin problemas la información recibida en el historial local y proporcione una imagen completa de la salud del paciente como si se hubiera obtenido localmente. Además, los datos de HCE anonimizados pueden alimentar directamente las herramientas de análisis de las autoridades sanitarias y los investigadores.

Todas estas definiciones de interoperabilidad entre sistemas de información sanitarios coinciden en plantear diferentes niveles de interoperabilidad, identificando el nivel semántico como el más complejo y avanzado. Los sistemas sanitarios se enfrentan a muchas barreras para lograr una implementación exitosa de sistemas semánticamente interoperables. Desde la falta de una biblioteca completa y exhaustiva de conocimientos médicos compartibles y procesables, hasta la dificultad para resolver las ambigüedades de las guías clínicas informatizadas (GCI). Además, los datos de los pacientes pueden residir en varios repositorios sin que haya un registro completo de todos los datos disponibles en una sola institución, lo que dificulta su uso clínico y supone un reto para los usos secundarios de la HCE o la integración de la misma con SADC [81]. Para lograr un intercambio interoperable de HCE, deben considerarse y definirse varios requisitos (técnicos, arquitectónicos, etc.). Estos requisitos se traducen formalmente en normas o estándares proporcionados por las

organizaciones adecuadas, como las organizaciones de desarrollo de estándares (ODE). En función de su ámbito de influencia, las ODE pueden ser internacionales, nacionales o regionales. Las ODE más relevantes para la normalización de la HCE a nivel mundial son independientes (a excepción de los comités técnicos), pero su desarrollo y resultados paralelos en campos de trabajo similares ha propiciado la necesidad de establecer relaciones de colaboración entre ellas. La comunicación y la cooperación entre los diferentes sistemas de HCE y sus componentes en un entorno complejo y altamente dinámico requiere la adopción de terminologías y ontologías comunes, servicios avanzados de seguridad, protección y privacidad y la separación entre los puntos de vista lógico y terminológico [82]. La existencia de una variedad tan amplia de requisitos implica la definición de normas especializadas, clasificadas según el ámbito específico al que se refieran. La incorporación de estos estándares y modelos en los nuevos diseños de sistemas de HCE permitiría dar un paso más en la utilización de los futuros sistemas de e-Salud, evitando la proliferación de prototipos no reutilizables.

2.8. Soluciones interoperables de integración de sistemas de HCE con resultados de SADC y datos enlazados abiertos

Las ventajas de combinar la HCE con resultados de SADC y contenidos procedentes de fuentes de información heterogéneas son bien conocidas, y han sido ampliamente descritas en la literatura. Existe la necesidad de reutilizar, además de la HCE, los resultados de SADC y los diccionarios de términos, a través de las fronteras de las organizaciones para minimizar los costes de desarrollo y mantenimiento [83]. En este contexto, el uso de sistemas de HCE estandarizados adquiere especial relevancia, dado que los SADC se alimentan en gran medida de datos de HCE, la forma en que se almacenan y transmiten dichos datos debe ser internacionalmente conocida y comúnmente aceptada por la comunidad de sistemas de información sanitaria (es decir, estandarizada). Una HCE estandarizada, permite alimentar SADC mediante abstracciones de HCE obtenidas a partir de transformaciones encadenadas de datos basadas en arquetipos, promoviendo además su reutilización en diversos escenarios [5, 84]. Esta misma idea es aplicable al cribado de datos en la elegibilidad de muestras para ensayos clínicos [6]. En ambos casos, una abstracción de la HCE proporciona una entrada adecuada para otro sistema mediante la transformación de instancias a través de una cadena de arquetipos de HCE. El uso de estándares es esencial en este contexto, pero se necesita algo más que estándares para abordar la interoperabilidad entre los sistemas de HCE y de SADC, debido a la gran variabilidad de SADC y a las particularidades de la HCE almacenada en diferentes organizaciones [85].

La literatura científica pone de manifiesto la relevancia de la web semántica aplicada a los sistemas de información sanitarios, en diversas líneas de investigación: metodologías de razonamiento/inferencia de la HCE, metodologías de consulta basadas en la semántica para la explotación de datos de HCE, aplicaciones para la validación de arquetipos/instancias, integraciones de sistemas de HCE con SADC, transformaciones de datos de HCE entre estándares de HCE, transformación de la HCE en formato de la web semántica, enriquecimiento de los datos de HCE con conocimiento semántico y otros usos secundarios de la HCE. Por ejemplo, es posible aplicar el razonamiento de hechos y reglas del lenguaje de reglas de la web semántica SWRL (Semantic Web Rule Language) sobre ontologías, para inferir detecciones tempranas de interacciones entre fármacos y sus mecanismos de interacción farmacocinética y farmacodinámica a gran escala [86]. O validar asociaciones entre fármacos y eventos en ontologías y repositorios de datos, analizando si una asociación establecida a partir de múltiples fuentes complementarias tiene más probabilidades de ser cierta [87]. Ambos casos pueden contribuir al enriquecimiento de la HCE, integrando el conocimiento resultante o la información derivada de cada trabajo en la HCE.

La transformación de instancias de datos de HCE normalizados entre estándares de HCE basados en el modelo dual (openEHR e ISO13606), mediante herramientas de la web semántica y los lenguajes RDF y OWL (Ontology Web Language), permite superar las barreras en la transferencia de datos de HCE entre organizaciones que cumplen con estos estándares [88]. Incluso, es posible implementar dichas transformaciones entre estándares de HCE no duales, por ejemplo, mediante el modelado de recursos HL7 FHIR con arquetipos ISO13606 [22]. No obstante, dichas transformaciones no introducen por sí solas contenidos adicionales potencialmente relevantes en la HCE, ni realizan razonamiento o inferencia sobre los datos de la misma. Es posible transformar instancias de HCE normalizada a RDF/OWL para su posterior razonamiento con SWRL manteniendo la estructura del MR de HCE en formato RDF/OWL y, a la vez, permitiendo explorar las características ontológicas de ese modelo mediante consultas semánticas [89]. Sin embargo, este enfoque tiene limitaciones debido a su abstracción de dominios específicos, por lo que las consultas semánticas son efectivas principalmente a nivel estructural, no a nivel de dominio del conocimiento. Para explotar los datos a nivel de dominio, es necesario diseñar reglas SWRL adaptadas a cada escenario. Un marco común para el enriquecimiento semántico de los datos clínicos mediante enlaces terminológicos, facilita establecer relaciones con terminologías/ontologías y transformar automáticamente datos clínicos del mundo real en individuos OWL adecuados. Haciendo uso de estos individuos OWL es posible realizar consultas semánticas a nivel de dominio del conocimiento clínico [90]. Estudios sobre la representación del modelo de elementos clínicos CEM (Clinical Elements Model) con RDF/OWL, demuestran su beneficio al aportar una definición semántica formal a las especificaciones del CEM. Dicha definición

semántica, facilita la conexión del CEM con varias tecnologías de la web semántica para añadir capacidades de autoría, razonamiento y consulta de los contenidos del CEM [91], permitiendo la armonización de modelos CEM con modelos clínicos en otras representaciones como por ejemplo el lenguaje de definición de arquetipos ADL (Archetype Definition Language). La idea de transformar la información clínica a formatos de la web semántica también se ha investigado aplicada a datos LOD, con el propósito de explotar las ventajas que ofrecen las consultas federadas a repositorios de datos LOD públicos en la nube [92]. La investigación en ciencia clínica exige métodos eficaces para acceder e integrar datos de fuentes de datos múltiples y heterogéneas de forma unificada. La web semántica permite la explotación integrada de recursos biomédicos, típicamente dispuestos a lo largo de múltiples fuentes heterogéneas, mediante la transformación e integración de datos biomédicos aplicando los principios de LOD [93]. Esta metodología facilita la generación de conjuntos de datos biomédicos abiertos e interoperables en formatos de la web semántica. El enriquecimiento de la HCE normalizada con información derivada de estos trabajos, requiere de la normalización de dicha información de acuerdo a estándares de HCE y de una integración adecuada con los datos iniciales de HCE.

Las tecnologías de la web semántica proporcionan mecanismos de consultas semánticas e inferencia, que aplicados a la HCE permiten la búsqueda de información basada en conceptos, superando los desafíos específicos de los métodos de búsqueda sintácticos en la HCE [94, 95]. Además, son útiles para validar de arquetipos openEHR e ISO13606. Mediante una representación RDF/OWL de un arquetipo, expresado en el lenguaje de definición de arquetipos ADL, es posible realizar razonamiento sobre su estructura y establecer la validez del mismo [96]. Se obtiene una validación mayor al incorporar información de dominio en el proceso de razonamiento, aunque requiere modelar también la información de dominio que representa el arquetipo en RDF/OWL [97]. La incorporación de información de dominio al proceso de razonamiento permite combinar eficazmente arquetipos y ontologías para otros fines, como la identificación de cohortes de pacientes, la elaboración de perfiles de fenotipos y la identificación de genes de enfermedades [16], [98].

La mayoría de los trabajos que combinan la HCE con contenidos procedentes de fuentes de información heterogéneas, se centran en la transformación de los datos de la HCE a una representación semántica, para realizar razonamientos e inferencias en usos secundarios de la HCE. Lezcano et al. [89] presenta una solución para el razonamiento de arquetipos clínicos utilizando ontologías OWL y expresiones SWRL. Argumentaron que los lenguajes de arquetipos actualmente disponibles no proporcionan un soporte directo para mapear a ontologías formales y luego explotar el razonamiento sobre el conocimiento clínico. Tradujeron las definiciones expresadas en ADL de openEHR a una representación formal de OWL, y luego aplicaron las reglas de SWRL a instancias de datos clínicos. Se han empleado

enfoques similares para mejorar los usos secundarios de la HCE. Tao et al. [91] introdujeron una representación OWL del modelo de elementos clínicos CEM, para realizar el razonamiento y la comprobación de la consistencia de los modelos derivados para los usos secundarios de la HCE. J. T. Fernández Breis et al. [16] aplicaron tecnologías de la web semántica para aprovechar los estándares de HCE para la identificación de cohortes de pacientes. Ambos trabajos establecieron enfoques para convertir los datos de HCE en una representación semántica para el razonamiento. Sin embargo, los datos de HCE expresados como representación semántica tienen un grado limitado de interoperabilidad en los sistemas informáticos sanitarios normalizados de acuerdo con arquitecturas de HCE basadas en modelo dual.

La conversión de los datos de HCE en una representación semántica ha sido ampliamente explorada para mejorar la interoperabilidad de la HCE. D. J. Odgers et al. [98] diseñaron un método para la extracción de datos de HCE utilizando datos enlazados. Transformaron una versión anonimizada de la base de datos STRIDE de Stanford [99] en un almacén de datos clínicos semánticos para realizar la selección de cohortes, el perfil de fenotipos y la identificación de genes de enfermedades. V. Kilintzis et al. [100] presentaron un marco de gestión de datos de telemedicina para apoyar los servicios de atención integrada a pacientes crónicos. Lo construyeron sobre HL7 y ontologías para obtener una representación semánticamente enriquecida de los datos de la HCE siguiendo los principios de los datos enlazados. Sin embargo, ningún trabajo del estado del arte actual aborda transformaciones en el sentido contrario, desde la representación de la web semántica a una representación formal de HCE. J. J. Cimino et al. [101] introdujeron Infobuttons, un método para acceder a recursos de conocimiento externos desde la HCE. Consiste en un conjunto de especificaciones que definen URLs parametrizadas con datos del extracto de HCE. Cuando el usuario selecciona uno de estos enlaces, se recupera información adicional de una fuente externa y se muestra al usuario como HTML. Aunque el enfoque de Infobuttons se ha incorporado como parte del estándar HL7, su principal limitación es la falta de interoperabilidad de esta solución con los estándares de HCE basados en el modelo dual. Asimismo, M. Alfano et al. [102] presentaron un método para obtener definiciones adicionales de términos clínicos. Parte de extractos de HCE y consulta un tesoro para proporcionar explicaciones de términos médicos a través de un servicio web. A. Chetta et al. [103] introdujeron una solución no estándar para enriquecer las HCE con contenidos adicionales del servicio de enfermería. Ayuda a las enfermeras a documentar y razonar sobre los datos de los pacientes y sobre la comprensión de los mismos por parte del médico. Las soluciones de M. Alfano et al. o de A. Chetta et al. ponen de manifiesto la importancia de enriquecer los datos de la HCE con información potencialmente relevante procedente de fuentes externas. Sin embargo, existe una falta de estandarización de la información adicional combinada con

los extractos de HCE, lo que conduce a un grado limitado de interoperabilidad de dicha información adicional en los sistemas informáticos sanitarios normalizados. La HCE aumentada que se presenta en esta tesis pretende mitigar esta limitación enriqueciendo los extractos de HCE basados en arquetipos con información adicional procedente de fuentes de la web semántica, expresada en términos de arquetipos de HCE.

Capítulo 3. Historia Clínica Electrónica aumentada con datos enlazados abiertos

3.1. Introducción

La HCE de un paciente se puede complementar con información adicional potencialmente relevante y útil en varios escenarios [104, 105], incluidos los usos secundarios de la HCE o la interacción con SADC. Este contenido útil puede obtenerse de fuentes de datos públicas como la nube de datos enlazados abiertos LODC (Linked Open Data Cloud) [106], construida con datos LOD [107], o de otras fuentes de información heterogéneas susceptibles de ser transformadas en una representación semántica basada en LOD. Por ejemplo, es posible obtener el conjunto de fármacos contraindicados para un paciente, considerando su medicación activa y sus trastornos diagnosticados, e incorporarlo a la HCE mediante su integración con fuentes de información de la web semántica.

Esta tesis introduce la HCE aumentada (o enriquecida), entendida como una HCE existente complementada con contenidos adicionales procedentes de fuentes de información externas. Estos contenidos son cualquier tipo de información y/o conocimiento que pueda ser de interés para su inclusión en la HCE. Existen varios enfoques para conseguir la HCE aumentada, que van desde la transformación de la información de fuentes externas en un formato de HCE normalizado, hasta la transformación de la HCE al formato de dichas fuentes. En ambos casos, garantizar una integración completa requiere transformaciones de datos para obtener una representación de ambos contenidos expresada en el mismo formato. Dichas transformaciones requieren, a su vez, identificar las correspondencias existentes entre ambas fuentes para garantizar la coherencia y dominio semántico de la información. El enriquecimiento de la HCE aplica tanto a cada instancia individual de HCE (es decir, extractos de HCE), como de forma generalizada a todas las instancias de un determinado modelo de HCE, aumentando la información del esquema de HCE normalizada al que pertenecen las instancias. En el primer caso, se enriquece la HCE individualmente para cada paciente (por ejemplo, incluyendo información adicional relacionada con la medicación de un paciente), mientras que en el segundo caso, el enriquecimiento de la HCE se aplica a todos los pacientes (por ejemplo, la inclusión de explicaciones adicionales sobre el significado de la sección "medicación activa"). No obstante, la HCE y las fuentes externas se encuentran en continua evolución, por lo que en ambos casos, el enriquecimiento de la HCE debe realizarse siempre que las fuentes involucradas cambien y podría ser necesario implementar varias versiones de una HCE aumentada para garantizar su validez a lo largo del tiempo.

En este capítulo, se describe una metodología para abordar la HCE aumentada a nivel de instancia, enriqueciendo extractos de HCE con contenidos de la web semántica y presentando un caso de uso para demostrar su viabilidad. Además, también se describe cómo esta metodología funciona a nivel de esquema de HCE, para enriquecer el significado de la información contenida en el arquetipo de HCE. Nótese que el concepto de HCE aumentada, tal y como se introduce aquí, comprende la inclusión de información procedente de fuentes de información externas y heterogéneas en la HCE, independientemente de la representación inicial de dicha información en sus fuentes originales. Aunque la construcción de la HCE aumentada requiere de fuentes de información externas representada en forma de datos LOD, la metodología presentada también es válida para obtener la HCE aumentada utilizando contenidos de otras fuentes heterogéneas y/o incluso de otras HCE, transformándolas previamente a una representación de datos enlazados. Específicamente, la metodología parte de un arquetipo de HCE y se basa en la identificación y definición formal de enlaces entre los elementos de dicho arquetipo y una fuente de información de la web semántica. El enlace, que debe diseñarse manualmente, se utiliza para identificar y recoger información adicional (en adelante, el "Contenido de Aumento") potencialmente relevante para el paciente. El Contenido de Aumento (CA) es normalizado como un extracto de HCE e integrado en la HCE original, dando lugar a un extracto de HCE aumentado resultante de la integración completa de ambas fuentes. Nótese que en este trabajo "extracto de HCE" e "instancia de arquetipo de HCE" se utilizan indistintamente para referirse a los datos de salud del paciente normalizados. Por último, este capítulo aborda como caso de uso la HCE aumentada del arquetipo de HCE resumida, publicado como parte del CMDIC. Un caso de uso en el que la HCE resumida original se enriquece con CA de dos fuentes de la web semántica diferentes: la representación LOD de la base de datos DrugBank [108] para obtener las interacciones entre medicamentos según la medicación activa del paciente, y la terminología de referencia del archivo nacional de fármacos NDFRT (National Drug File-Reference Terminology) [26] para recuperar información de tratamientos y efectos relacionados con las enfermedades específicas (problemas o episodios activos) del paciente.

3.2. Planteamiento

Nuestro trabajo de investigación se centra en las arquitecturas de HCE de modelo dual, que consisten en definir por separado el esquema de HCE (arquetipos de HCE) de los datos de HCE (instancias o extractos de HCE). En concreto, el modelo dual define dos modelos diferentes: un MR genérico diseñado para representar las propiedades y estructuras más básicas de cualquier HCE, ejemplos de ello son openEHR e ISO13606; y un MA. Los arquetipos son definiciones detalladas de conceptos clínicos en forma de combinaciones estructuradas y restringidas de las entidades del MR [4]. Los arquetipos se utilizan para

modelar conceptos en la HCE, y los extractos de HCE son instancias de arquetipos que contienen los datos de HCE del paciente.

Este capítulo introduce una metodología para enriquecer una HCE basada en un modelo dual con información potencialmente relevante de la web semántica. La web semántica incorpora a la web metadatos semánticos procedentes de ontologías y terminologías. Define identificadores de recursos uniformes o URI (Uniform Resource Identifier) que apuntan no solo a documentos sino también a datos, permitiendo la categorización de cada dato en la web semántica con valores codificados de terminologías y ontologías. Además, la web semántica establece mecanismos para definir, también mediante URI, la relación semántica entre los datos. La web semántica proporciona a las máquinas la capacidad de interpretar y hacer inferencias sobre los datos [109]. Los lenguajes de metadatos, como RDF u OWL, permiten indexar los recursos de información de la web con sus representaciones de conocimiento y almacenarlos en documentos. La web semántica, además, introduce mecanismos para consultar y recuperar datos como el lenguaje de consulta SPARQL (SPARQL Protocol And RDF Query Language), y para realizar razonamientos sobre los datos como el lenguaje SWRL.

Enriquecer los extractos de HCE con datos procedentes de fuentes de información externas heterogéneas es uno de los principales objetivos de esta tesis. Para lograr una solución ampliamente aplicable, los puntos de partida son (i) instancias de HCE normalizadas según un conjunto de arquetipos de HCE (es decir, HCE normalizadas de acuerdo al modelo dual con arquetipos ISO13606 u openEHR), y (ii) contenido de aumento expresado en formato de la web semántica (estructurado y almacenado como conjuntos de datos RDF). Otro conjunto de arquetipos de HCE, modelados específicamente para representar el contenido de aumento como extractos de HCE normalizados (en adelante, arquetipos de aumento), permiten homogeneizar el formato del contenido de aumento y el extracto de HCE inicial en una misma representación normalizada. Un último arquetipo de HCE llamado arquetipo aumentado, que comprende el arquetipo inicial y los arquetipos de aumento, permite una integración completa del extracto de HCE inicial y del contenido de aumento. Nótese que el arquetipo aumentado puede contener uno o más arquetipos de aumento. El arquetipo aumentado puede considerarse como una especialización del arquetipo inicial, pues se trata de una extensión de este último que incluye información adicional (las instancias de los arquetipos de aumento). En esta tesis, el proceso para definir una HCE aumentada se basa en estándares de HCE, arquetipos y mapeos entre el contenido de aumento y arquetipos de HCE. Formalmente, este proceso se define en términos de dos problemas de transformaciones de datos: normalizar los contenidos de aumento en extractos de HCE (transformación de datos 1), y combinarlos con el extracto de HCE inicial (transformación de datos 2) dando lugar al extracto de HCE aumentado. Cada transformación de datos se expresa, a nivel de esquema, mediante una

tupla de 3 elementos $\langle [EF], ED, M \rangle$. Obsérvese que esta tupla se utiliza para especificar una operación de transformación de datos, y no debe confundirse con tuplas RDF. En cada tupla, $[EF]$ representa uno o varios esquemas fuente, ED un esquema de destino y M un conjunto de reglas de mapeo entre $[EF]$ y ED . Por otra parte, a nivel de instancias la transformación de datos se representa con la transformación T , que aplica las reglas de mapeo M para transformar instancias de datos de los esquemas fuente $[EF]$ en instancias de datos del esquema destino ED .

El primer problema de transformación de datos consiste en normalizar el contenido de aumento. Esto se define a nivel de esquema por la tupla $\langle [EF_1], ED_1, M_1 \rangle$, donde EF_1 , el esquema de datos para contenidos de la web semántica, es una representación canónica del contenido de aumento expresado en formato de la web semántica. ED_1 , el arquetipo de aumento, es una forma canónica del contenido de aumento expresado en arquetipos de HCE. Por último, M_1 es la especificación de las relaciones entre los elementos de EF_1 y los elementos de ED_1 . A nivel de datos, la transformación T_1 es una implementación de M_1 destinada a normalizar contenidos de aumento en extractos de HCE, convirtiendo instancias de EF_1 en instancias de ED_1 .

La segunda operación de transformación de datos consiste en combinar el extracto inicial de HCE con los contenidos de aumento normalizados, dando lugar al extracto de HCE aumentada. La tupla $\langle [EF_{2.1}, EF_{2.2}], ED_2, M_2 \rangle$ define esta transformación de datos a nivel de esquema, donde $EF_{2.1}$ es el arquetipo inicial y $EF_{2.2}$ es el arquetipo de aumento. Los dos esquemas de origen son representaciones canónicas del extracto inicial de HCE y del contenido de aumento normalizado, respectivamente. El esquema de destino ED_2 , el arquetipo aumentado, es la representación canónica del extracto de HCE aumentada. M_2 representa las especificaciones de mapeo entre los elementos de $[EF_{2.1}, EF_{2.2}]$ y los elementos de ED_2 . A nivel de datos, la transformación T_2 es una implementación de M_2 destinada a generar extractos de HCE aumentados convirtiendo instancias de $[EF_{2.1}, EF_{2.2}]$ en instancias de ED_2 . Nótese que ED_1 y $EF_{2.2}$ son ambos el mismo esquema (arquetipo de aumento), lo que implica que el resultado de la transformación T_1 es una de las entradas utilizadas en la transformación T_2 .

En el caso de uso presentado en esta tesis (la HCE resumida aumentada) se emplean el esquema del W3C para resultados de consultas SPARQL [110] como forma canónica de los contenidos de la web semántica (EF_1), los arquetipos ISO13606 como representación canónica de los extractos de HCE ($EF_{2.1}, EF_{2.2}, ED_1$ y ED_2), y el lenguaje de mapeo declarativo de alto nivel de la plataforma LinkEHR [17, 111] para definir especificaciones de mapeo (M_1 y M_2) a nivel de esquema, y para generar transformaciones de datos (T_1 y T_2) a nivel de datos. La figura 8 resume los procesos de transformaciones de datos descritos.

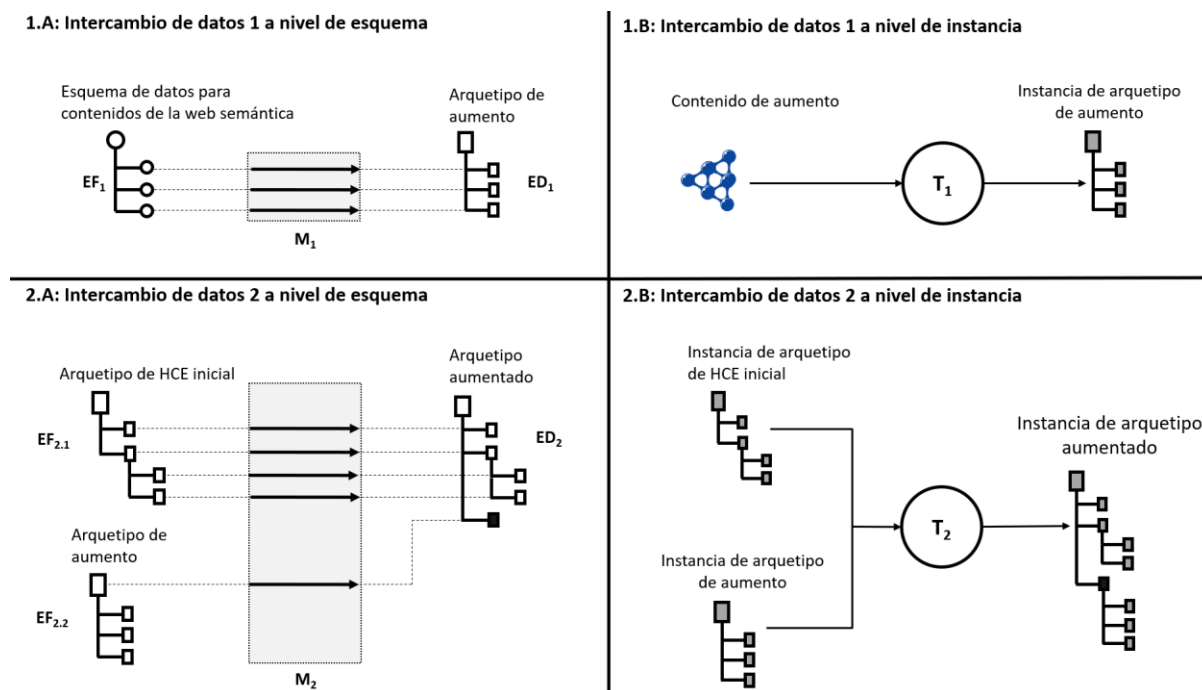


Figura 8. Operaciones de intercambio de datos para construir extractos de HCE aumentados. (1.A) A nivel de esquemas y (1.B) a nivel de datos representan las operaciones para normalizar el contenido de aumento como un extracto de HCE, mientras que (2.A) a nivel de esquemas y (2.B) a nivel de datos muestran las operaciones para transformar el extracto de HCE inicial junto con el contenido de aumento normalizado en un extracto de HCE aumentado. Nótese la diferencia entre M, que es la especificación de relaciones o mapeos entre esquemas de datos, y T, que representa la implementación de transformaciones de instancias de datos de acuerdo a la especificación definida por M.

Cada arquetipo de HCE puede contener varios ítems susceptibles de ser complementados con información adicional procedente de fuentes externas (por ejemplo, valores codificados que identifiquen un diagnóstico). Estos elementos son las entidades de interés (EDI en adelante). Las EDI están representadas por una ruta desde la raíz del arquetipo de HCE hasta el nodo de la EDI, y constituyen las partes de un arquetipo de HCE que se desea aumentar. Además, las EDI permiten identificar los contenidos de las fuentes externas a considerar como contenido de aumento para enriquecer la HCE. Las EDI son el vínculo entre la HCE y el contenido de aumento, y son el punto de partida para explorar las fuentes externas en busca de contenido relacionado. Aquellos contenidos potencialmente relevantes en el contexto del arquetipo inicial se emplean como contenido de aumento.

El diseño y creación de un arquetipo aumentado para un determinado arquetipo de HCE, utilizando la web semántica como fuente externa, requiere la realización de las siguientes tareas y modelados a nivel de esquema (figura 9):

1. Identificar las EDI en el arquetipo inicial. Estas entidades suelen ser valores codificados y deben seleccionarse manualmente. Alternativamente, los valores de texto libre pueden ser utilizados como EDI siempre que estos valores sean traducidos a valores codificados (por ejemplo, aplicando métodos de procesamiento de lenguaje natural o consultando repositorios terminológicos).

2. Para cada EDI, identificar un contenido de aumento potencialmente relevante de una fuente externa. Para ello es necesario seleccionar una fuente de información relacionada con los valores codificados de la EDI. Dicha fuente debe incluir elementos de datos que representen los conceptos de la EDI (por ejemplo, un conjunto de medicamentos o un conjunto de problemas o episodios activos), y debe contener información potencialmente relevante para enriquecer el arquetipo inicial. El contenido de aumento se obtiene de esta fuente.
3. Garantizar la viabilidad de la vinculación de la HCE con el contenido de aumento, mediante la búsqueda de mapeos entre las EDI y los elementos de datos equivalentes (en adelante, EDE) en la fuente externa. Normalmente, diferentes terminologías utilizan códigos diferentes para representar el mismo concepto. Este desajuste del código del concepto entre terminologías puede superarse empleando servicios de mapeo terminológico. Además, los descriptores VoID (Vocabulary of Interlinked Datasets) [112] pueden ser útiles para encontrar estos mapeos porque permiten identificar los conjuntos de datos públicos de la web semántica (como los disponibles en la LODC) que contienen EDE codificados con otras terminologías. Estos descriptores son vocabularios públicos de conjuntos de datos interconectados, que contienen metainformación sobre los repositorios de datos LOD en la web semántica.
4. Construir el enlace entre la HCE y el contenido de aumento. La vinculación debe enlazar las EDI de la HCE inicial con los EDE de la fuente externa, preservando el significado de las EDI en el contexto del contenido de aumento, que está expresado con términos de la fuente externa. El enlace se define mediante tripletas de RDF (en adelante, las tripletas de enlace EDI). Aunque el método presentado permite aumentar la HCE con múltiples fuentes de información heterogéneas, es necesaria una conversión previa de las fuentes externas a formato de la web semántica en caso de que estas se encuentren en un formato distinto. La construcción de las tripletas de enlace EDI puede automatizarse con un procedimiento que, partiendo de los valores codificados de la EDI y del nombre de las terminologías implicadas, utiliza un servicio de mapeo terminológico y, opcionalmente, la información recogida de los descriptores VoID, para identificar los EDE y con ellos generar las tripletas de enlace EDI.
5. Obtener el contenido de aumento de la fuente externa mediante las tripletas de enlace EDI. Una consulta SPARQL combina las tripletas del conjunto de datos de la fuente externa y las tripletas de enlace EDI para obtener el contenido de aumento. Dicha consulta debe ser diseñada manualmente para garantizar que se obtiene contenido de aumento relevante en el contexto de las EDI y el arquetipo inicial. Finalmente, el resultado de la consulta proporciona el contenido de aumento en forma de tripletas RDF.

6. Modelado y mapeo del arquetipo de aumento. El propósito de este arquetipo es normalizar el contenido de aumento como un extracto de HCE. El arquetipo de aumento debe contener al menos un elemento por cada dato incluido en el contenido de aumento. El mapeo del esquema de datos del contenido de aumento y el arquetipo de aumento se define estableciendo la relación entre las entidades de ambos esquemas. En su expresión más sencilla, este mapeo consiste en la relación 1:1 de las entidades de ambos esquemas. No obstante, es posible definir sofisticados mapeos que involucren transformaciones más complejas (por ejemplo, agregación de datos) entre las entidades de ambos esquemas. Nótese que si se aplica la versión de mapeo más simplificada (es decir, relación 1:1), es posible modelar automáticamente el arquetipo de aumento definiendo un CLUSTER como entidad raíz. Dicho CLUSTER contiene un elemento por cada entidad del contenido de aumento. De igual manera, se puede automatizar la especificación del mapeo para este CLUSTER, dado que se trata de una relación 1:1, sin ambigüedad, entre las entidades de ambos esquemas. Este paso podría realizarse una vez identificado el contenido de aumento (paso 2), en paralelo con los pasos 3, 4 y 5.
7. Modelado y mapeo del arquetipo aumentado. Este arquetipo combina el arquetipo inicial y el/los arquetipos de aumento. Al igual que en el paso anterior, deben definirse las especificaciones de mapeo entre los esquemas de origen (arquetipo inicial y arquetipo de aumento) y el esquema de destino (arquetipo aumentado). Este paso debe darse una vez que se haya completado el paso 6. El arquetipo aumentado es el resultado de especializar el arquetipo inicial con uno o más SLOT para incluir los arquetipos de aumento en lugares específicos. Aunque esta tesis más adelante introduce un algoritmo para sugerir posibles ubicaciones adecuadas para dichos SLOT, las ubicaciones finales deben seleccionarse manualmente para supervisar la selección automática y preservar el significado del arquetipo inicial. Una vez elegida una ubicación para cada SLOT, el arquetipo aumentado puede generarse automáticamente. Además, la definición del mapeo puede automatizarse tras determinar las posiciones de los SLOT.

Hay que tener en cuenta las siguientes consideraciones. En primer lugar, los pasos del 2 al 6 definen el arquetipo de aumento, mientras que los pasos del 1 al 7 describen todo el proceso para obtener el arquetipo aumentado. Por otra parte, un arquetipo aumentado puede incluir diferentes arquetipos de aumento, al menos uno por cada EDI identificada a nivel de esquema en el arquetipo inicial. En tal caso, el proceso implicaría el diseño de varios arquetipos de aumento. Existen diferencias significativas entre la definición de un arquetipo aumentado y la generación de instancias de un arquetipo aumentado. La definición es una tarea de modelado

y diseño. Se realiza a nivel de esquema y requiere la intervención manual en algunos puntos. En cambio, la generación de instancias se produce a nivel de datos y se efectúa de forma totalmente automática. Así, una vez definida una HCE aumentada, esta sirve para enriquecer automáticamente cualquier extracto de HCE que se ajuste al arquetipo inicial. Por último, la cardinalidad de las EDI en el arquetipo inicial determina la cardinalidad del elemento raíz (normalmente un CLUSTER) del arquetipo de aumento. Una cardinalidad múltiple de una EDI puede dar lugar a varias tripletas de enlace EDI (una por cada valor codificado que el extracto inicial de HCE contenga en dicha EDI). Por lo tanto, pueden encontrarse varios conjuntos de contenido de aumento que deben encajar en la misma instancia del arquetipo de aumento. Así pues, la cardinalidad del elemento raíz del arquetipo de aumento debe ser mayor o igual que la cardinalidad de la EDI que constituye el enlace con este arquetipo de aumento.

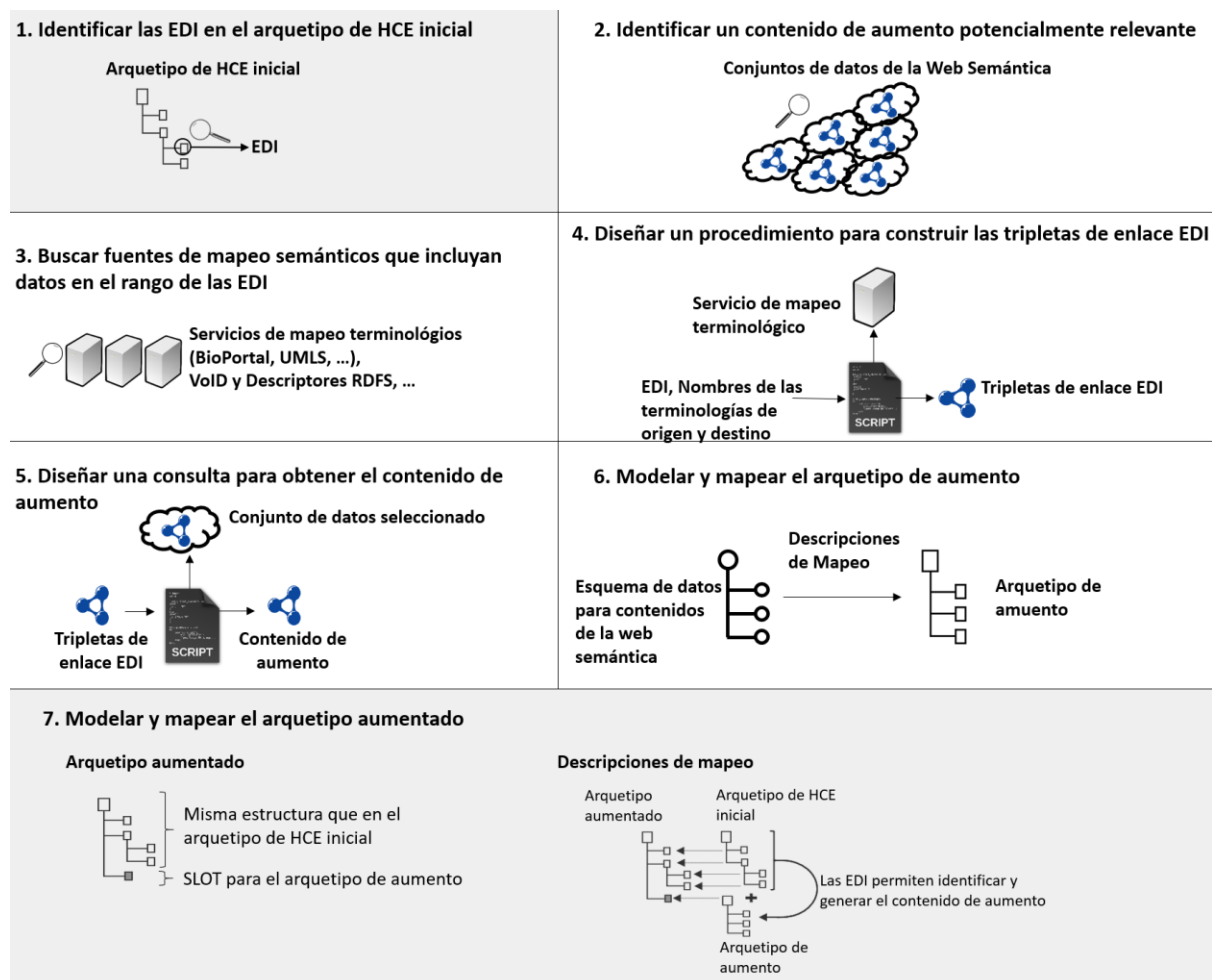


Figura 9. Metodología para definir el arquetipo de aumento (pasos 2 a 6) y el proceso completo para obtener el arquetipo aumentado (pasos 1 a 7).

3.3. Metodología

A continuación, se detalla la metodología para diseñar una HCE aumentada de acuerdo con el planteamiento de esta tesis. En aras de la claridad, algunos de los pasos de la figura 9 se

fusionan en la explicación. También se incluye un ejemplo paso a paso para una mejor comprensión de la metodología. El ejemplo muestra cómo construir el arquetipo de medicación aumentado, partiendo del arquetipo de medicación y añadiendo información sobre las interacciones entre medicamentos como contenido de aumento. Como se ha mencionado anteriormente, el punto de partida es un arquetipo de HCE.

3.3.1. Identificación del contenido de aumento para un arquetipo

El diseño de un arquetipo aumentado requiere (i) identificar las EDI en el arquetipo de HCE, (ii) identificar las fuentes de información externas potencialmente interesantes y (iii) encontrar los EDE entre ellas (pasos 1, 2 y 3 de la figura 9). Como se ha dicho, normalmente las EDI son valores codificados (por ejemplo, diagnósticos, procedimientos y medicamentos), que pueden vincularse a fuentes de información externas para enriquecer el extracto inicial de la HCE con información o conocimientos adicionales. Estas fuentes de información externa pueden ser conjuntos de datos de la web semántica, u otras fuentes de información susceptibles de tener una representación de sus datos en formatos de la web semántica mediante transformaciones previas. Opcionalmente, los descriptores VoID pueden ayudar a identificar conjuntos de datos LOD públicos explotables mediante consultas SPARQL. Una vez identificada la representación en formato web semántica de la fuente de información externa, se deben localizar conjuntos de datos de la misma potencialmente interesantes para complementar la información que contienen las EDI. Los servicios de mapeo terminológicos sirven para identificar equivalencias entre valores codificados de terminologías distintas. Consultando a uno de estos servicios usando la información de las EDI, se obtienen los EDE de la terminología destino (utilizada por la fuente de información externa).

Ejemplo: Para construir el arquetipo de medicación aumentado, las EDI son enriquecidas con información adicional sobre las interacciones entre medicamentos. Como punto de partida, el arquetipo inicial es el arquetipo de HCE de medicación CEN-EN13606-ENTRY.Medicacion.v1, publicado como parte del CMDIC. Las EDI seleccionadas son los elementos de datos que contienen la medicación activa (localizados en el arquetipo en la ruta de acceso `/items[archetype_id="at0005"]/parts[archetype_id="at0010"]/value`). En este caso, las EDI se definen como tipo de datos CD⁴, que contienen valores codificados como por ejemplo códigos SNOMED-CT [24]. A partir de un subconjunto arbitrario de códigos de medicamentos SNOMED-CT, se consulta el servicio de mapeo terminológico BioPortal [113, 114] para identificar los correspondientes EDE en otras terminologías/ontologías. Para cada

⁴ Descriptor de concepto. Tipo de datos de la norma ISO13606 para registrar conceptos codificados (típicamente códigos terminológicos).

elemento de datos del subconjunto muestral, el servicio devuelve una lista de EDE de otras fuentes y sus URL (por ejemplo, la URL de un elemento de la base de datos DrugBank: <http://www.drugbank.ca/drugs/DB00099>). Normalmente, las URL pertenecen a la organización que creó la terminología, que a menudo no proporciona un conjunto de datos público para explotar la terminología a través de un lenguaje de consulta semántico. En el caso del repositorio DrugBank, la organización proporciona la base de datos en formato relacional en vez de en formato RDF. En esta situación, los mecanismos de descripción disponibles en la LODC, como los descriptores VoID y/o las propiedades RDFS, son útiles para encontrar conjuntos de datos semánticos que contengan la terminología requerida. Una consulta SPARQL a los descriptores VoID sirve para identificar una representación semántica de la base de datos DrugBank disponible públicamente (por ejemplo, <http://bio2rdf.org/drug-bank:DB00099> es un elemento de DrugBank disponible en la LODC "bio2rdf"). De esta manera se obtiene una representación del conjunto de datos de DrugBank en formato RDF. En caso de no poderse obtener dicha representación públicamente, es necesario realizar una transformación previa de la base de datos publicada por la organización a su representación correspondiente en formato RDF.

3.3.2. Transformación de las EDI en una representación semántica

Se define una relación formal entre las EDI y sus EDE en el conjunto de datos de destino. Esta relación se expresa a través de las tripletas de enlace EDI, que son tripletas RDF compuestas por los valores codificados de las EDI como sujeto y los EDE como objeto (paso 4 de la figura 9). Estas tripletas se construyen utilizando predicados de ontologías de alto nivel, destinados a describir las relaciones entre diferentes ontologías y/o terminologías. La elección del predicado varía según el propósito del contenido de aumento y su relación con la EDI. Normalmente, predicados como `match` o `exactMatch` de la ontología SKOS [115] son apropiados para definir relaciones entre valores codificados de diferentes terminologías. La ontología SKOS introduce varios predicados útiles para definir relaciones de equivalencia, en distintos grados, entre valores codificados EDI y EDE: `mappingRelation`, `closeMatch`, `exactMatch`, `broadMatch`, `narrowMatch` y `relatedMatch`. Sin embargo, si se trata de una relación exacta entre valores codificados (es decir, representan exactamente el mismo concepto semántico), se recomienda utilizar `exactMatch` porque define con mayor precisión dicha relación de equivalencia.

En tiempo de ejecución, un procedimiento automático puede generar las tripletas de enlace EDI a partir de los elementos obtenidos en la sección 3.3.1: el extracto inicial de HCE, los nombres de las terminologías implicadas y la URL de un conjunto de datos semántico

existente que contenga el contenido de aumento. Para cada paso del proceso se requieren entradas concretas:

1. Identificar los valores codificados de la EDI (sujetos en las tripletas de enlace EDI).

Entradas necesarias:

- 1.1. Extracto inicial de la HCE
- 1.2. Ruta de acceso a la EDI en el arquetipo inicial

2. Establecer el predicado de las tripletas de enlace EDI. Entradas:

- 2.1. La URL del predicado (por ejemplo, skos:exactMatch)

3. Obtener los EDE en el conjunto de datos de destino (objetos en las tripletas de enlace EDI). Entradas:

- 3.1. El nombre de la terminología/ontología de origen empleada por los valores codificados de la EDI
- 3.2. La URL de la terminología/ontología de destino devuelta por el servicio de mapeo terminológico.

Ambas entradas (3.1 y 3.2) se utilizan, junto con los valores codificados de las EDI (paso 1), para consultar el servicio de mapeo terminológico y obtener los EDE de la terminología de destino.

4. Formatear los EDE (paso 3) con la URL de un conjunto de datos semántico existente.

Entrada:

- 4.1. La URL base de los elementos de datos en el conjunto de datos LOD de destino.
Esto permite utilizar los EDE obtenidos del servicio de mapeo terminológico como objetos en las tripletas de enlace EDI. De lo contrario, los EDE devueltos por el servicio de mapeo podrían no ser explotables mediante SPARQL.

A partir de estas entradas, se implementa un procedimiento que extrae los valores codificados de las EDI, consulta el servicio de mapeo terminológico para obtener los EDE, y construye las tripletas de enlace EDI con el predicado proporcionado. Finalmente, el algoritmo devuelve las tripletas de enlace EDI como resultado. Obsérvese que, para la obtención de un contenido de aumento determinado, el único argumento que varía para el enriquecimiento de distintas HCE es el extracto inicial de la HCE (1.1). El resto de los argumentos deben identificarse una primera vez en la etapa de diseño y permanecen válidos para todas las ejecuciones. Del mismo modo, el procedimiento descrito para construir las tripletas de enlace EDI solo debe

implementarse una vez y es válido para todas las ejecuciones. El ejemplo muestra cómo funciona este procedimiento.

Ejemplo: Siguiendo con el ejemplo introducido en la sección 3.3.1, el siguiente paso es desarrollar el algoritmo descrito en la sección 3.3.2 para construir las tripletas de enlace EDI. En el ejemplo, el algoritmo toma las siguientes entradas:

1.1. Un extracto de HCE basado en el arquetipo de HCE de medicación, que tiene un valor codificado para el medicamento Filgrastim (código SNOMED-CT 386948008) en la EDI "medicación activa".

1.2. La ruta de acceso a la EDI (es decir, el valor codificado del medicamento) en el arquetipo de HCE de medicación: /items[archetype_id="at0005"]/parts[archetype_id="at0010"]/value.

2.1. El predicado skos:exactMatch, cuya URL completa es <https://www.w3.org/2009/08/skos-reference/skos.html#exactMatch> [116].

3.1. El nombre de la terminología fuente que se utilizará en el servicio de mapeo terminológico/ontológico: SNOMEDCT.

3.2. El nombre del conjunto de datos diana o de destino para construir las tripletas de enlace EDI: DrugBank.

4.1. La URL base <http://bio2rdf.org/drugbank>: [117] que apunta a una representación LOD del conjunto de datos DrugBank.

Tomando estos argumentos de entrada, se generan las tripletas de enlace EDI realizando los pasos:

I. Extracción del valor codificado EDI (386948008) del argumento 1.1 utilizando la ruta del argumento 1.2.

II. Consulta al servicio de mapeo terminológico por el valor codificado EDI (386948008) y el nombre de la terminología fuente (entrada 3.1. SNOMEDCT): <http://data.bioontology.org/ontologies/SNOMEDCT/classes/http%3A%2F%2Fpurl.bioontology.org%2Fontology%2FSNOMEDCT%2F386948008/mappings?apikey=<apikey>> [114].

III. Obtención del EDE en el conjunto de datos de destino a partir de la respuesta del servicio de mapeo terminológico, filtrando por el nombre del conjunto de datos de destino (entrada 3.2. DrugBank):

@id: <http://www.drugbank.ca/drugs/DB00099> [25]

La URL <http://www.drugbank.ca> no proporciona una representación semántica del conjunto de datos de DrugBank. Se requiere un paso más (paso IV) para convertir este elemento de datos en una URL que pertenezca a un conjunto de datos LOD existente y explotable mediante consultas SPARQL.

IV. Utilizando el prefijo de la URL proporcionado en el argumento 4.1 (<http://bio2rdf.org/drugbank:>), más el valor codificado del EDE de la terminología de destino (DB00099). El resultado es una URL perteneciente a un conjunto de datos LOD existente que puede utilizarse en una consulta SPARQL:
<http://bio2rdf.org/drugbank:DB00099>

V. Construir las tripletas de enlace EDI utilizando el predicado del argumento 2.1 (`skos:exactMatch`):

`<http://purl.bioontology.org/ontology/SNOMEDCT/386948008>`

`<https://www.w3.org/2009/08/skos-reference/skos.html#exactMatch>`

`<http://bio2rdf.org/drugbank:DB00099>`

El predicado `skos:exactMatch` sirve para generar tripletas de enlace EDI de la forma `<SnomedDrug> skos:exactMatch <DrugBankDrug>`. Los valores de los sujetos y objetos de estas tripletas se obtienen en tiempo de ejecución. En concreto, `<SnomedDrug>` se define por el valor de la EDI en la instancia del arquetipo inicial, mientras que el EDE `<DrugBankDrug>` es un término del conjunto de datos DrugBank LOD equivalente a `<SnomedDrug>`.

3.3.3. Consulta de conjuntos de datos de la web semántica

El siguiente paso es construir una consulta SPARQL para recuperar el contenido de aumento. El grafo de la consulta está formado por el conjunto de datos de destino más las tripletas de enlace EDI (figura 10). La consulta se parametriza con las tripletas de enlace EDI, lo que permite buscar el contenido de aumento en el conjunto de datos de destino utilizando los valores codificados EDI como criterios de consulta. Por otro lado, el conjunto de datos de destino se incluye a través de un servicio (es decir, utilizando la cláusula `SERVICE` de las consultas SPARQL federadas). En tiempo de ejecución, este enfoque conduce a la identificación y obtención del contenido de aumento para un extracto de HCE dado (paso 5 de la figura 9), solicitando solo los elementos de datos del grafo de consulta que involucran a los predicados de las tripletas de enlace EDI. Aunque la fuente (terminología con la que están codificados los valores de las EDI) permanece fuera del grafo de consulta, es posible identificar el contenido de aumento en la cláusula `WHERE`, utilizando las tripletas de enlace

EDI como filtro. Por último, la cláusula SELECT establece qué subconjunto del conjunto de datos de destino constituye el contenido de aumento. Los resultados de la consulta SPARQL se formatean según la representación canónica de contenidos de la web semántica establecida por el esquema del W3C para resultados de consultas SPARQL.

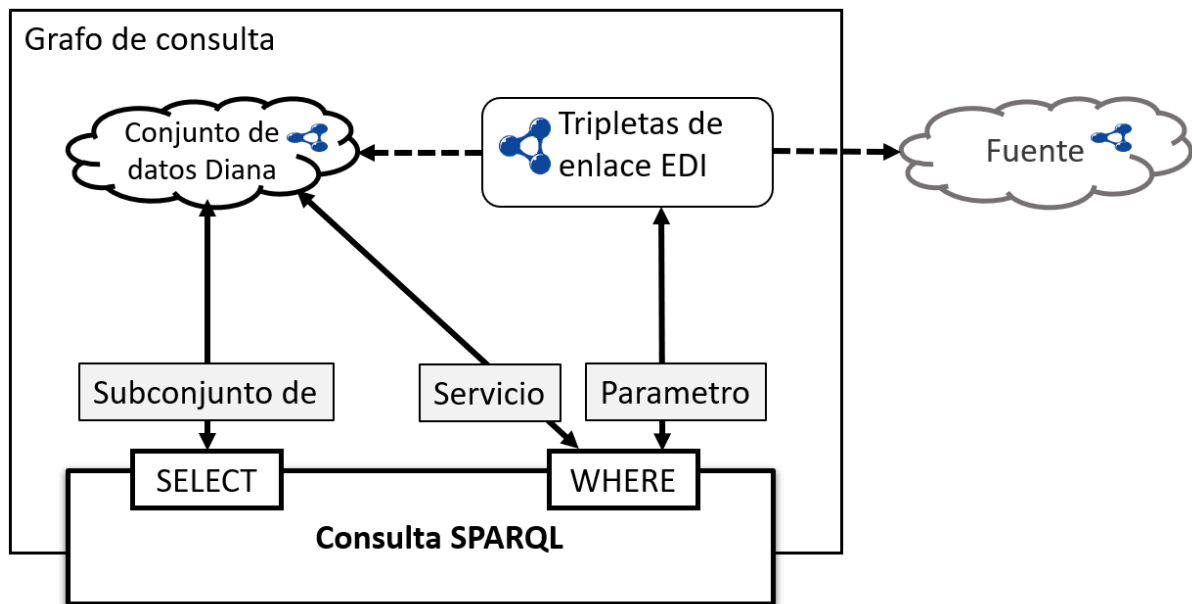


Figura 10. La consulta SPARQL y el grafo de consulta para obtener el contenido de aumento.

Ejemplo: El ejemplo ilustra una consulta SPARQL para obtener el contenido de aumento, consistente en información sobre interacciones entre fármacos de la medicación activa en el extracto de HCE (instancia del arquetipo de medicación). El contenido de aumento incluye descripciones de las interacciones, nombres y códigos de los fármacos que interactúan. La figura 11 muestra la consulta y el grafo de consulta de este ejemplo. El grafo de consulta consiste en el conjunto de datos del DrugBank (es decir, el conjunto de datos diana o de destino) más la tripleta de enlace EDI resultante del ejemplo de la sección 3.3.2. La primera parte de la consulta en la cláusula WHERE (?snomedDrugCode skos:exactMatch ?sourceDrugCode.) hace referencia a las tripletas de enlace EDI. La segunda parte de la consulta, la cláusula SERVICE, especifica cómo las tripletas de enlace EDI filtran el conjunto de datos diana. Por último, la cláusula SELECT especifica los elementos de datos del conjunto de datos diana que constituyen el contenido de aumento (?targetDrugName ?targetDrugCode ?interactionDescription). En tiempo de ejecución, en la tripleta de enlace EDI se establece el valor de ?sourceDrugCode como DB00099 (Filgrastim). A continuación, las tripletas de la cláusula SERVICE consultan el conjunto de datos DrugBank para obtener las descripciones de las interacciones entre fármacos

(?interactionDescription) relacionadas con el Filgrastim, además de los nombres (?targetDrugName) y códigos (?targetDrugCode) de los fármacos que interactúan.

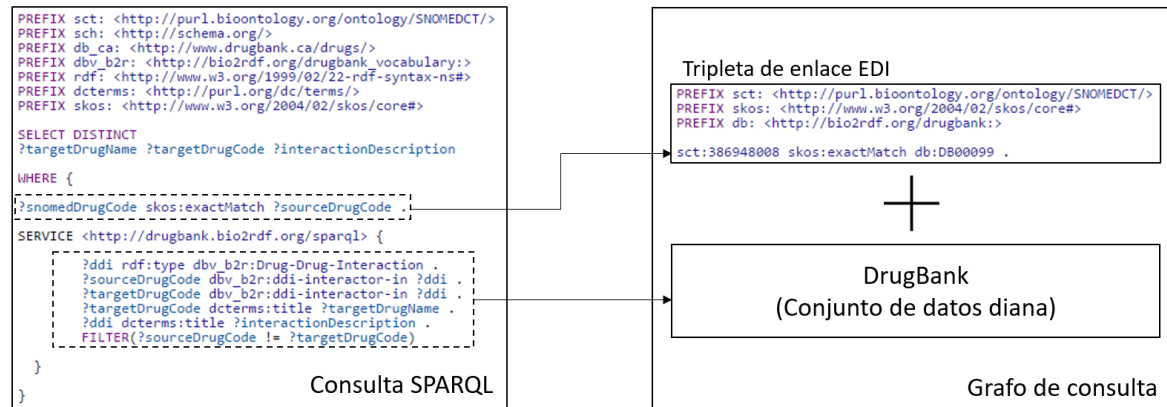


Figura 11. La consulta SPARQL y el grafo de consulta del ejemplo de la sección 3.3.3.

La ejecución de esta consulta SPARQL proporciona las interacciones farmacológicas del Filgrastim. La figura 12 muestra los resultados formateados según el esquema del W3C para resultados de consultas SPARQL. Obsérvese que las variables de la cláusula SELECT (figura 11) aparecen en el contenido del aumento (figura 12 y tabla 4).

```
<sparql xmlns="http://www.w3.org/2005/sparql-results#" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.w3.org/2001/sw/DataAccess/rf1/result2.xsd">
  <head>
    <variable name="targetDrugName"/>
    <variable name="targetDrugCode"/>
    <variable name="interactionDescription"/>
  </head>
  <results distinct="false" ordered="true">
    <result>
      <binding name="targetDrugName"><literal xml:lang="en">Bleomycin</literal></binding>
      <binding name="targetDrugCode"><uri>http://bio2rdf.org/drugbank:DB00290</uri></binding>
      <binding name="interactionDescription"><literal xml:lang="en">DDI between Filgrastim and Bleomycin
        - Filgrastim may enhance the adverse/toxic effect of Bleomycin.
        Specifically, the risk for pulmonary toxicity may be increased.</literal></binding>
    </result>
    <result>
      <binding name="targetDrugName"><literal xml:lang="en">Cyclophosphamide</literal></binding>
      <binding name="targetDrugCode"><uri>http://bio2rdf.org/drugbank:DB00531</uri></binding>
      <binding name="interactionDescription"><literal xml:lang="en">DDI between Filgrastim and Cyclophosphamide
        - Filgrastim may enhance the adverse/toxic effect of Cyclophosphamide.
        Specifically, the risk of pulmonary toxicity may be enhanced.</literal></binding>
    </result>
    <result>
      <binding name="targetDrugName"><literal xml:lang="en">Topotecan</literal></binding>
      <binding name="targetDrugCode"><uri>http://bio2rdf.org/drugbank:DB01030</uri></binding>
      <binding name="interactionDescription"><literal xml:lang="en">DDI between Filgrastim and Topotecan
        - Filgrastim may enhance the adverse/toxic effect of Topotecan.</literal></binding>
    </result>
  </results>
</sparql>
```

Figura 12. Resultados de la consulta SPARQL del ejemplo de la sección 3.3.3 formateados en XML, según el esquema canónico del W3C para los resultados de las consultas.

targetDrugName	targetDrugCode	interactionDescription
Bleomycin	http://bio2rdf.org/drugbank:DB00290	DDI between Filgrastim and Bleomycin - Filgrastim may enhance the adverse/toxic effect of Bleomycin. Specifically, the risk for pulmonary toxicity may be increased.
Cyclophosphamide	http://bio2rdf.org/drugbank:DB00531	DDI between Filgrastim and Cyclophosphamide - Filgrastim may enhance the adverse/toxic effect of Cyclophosphamide. Specifically, the risk of pulmonary toxicity may be enhanced.
Topotecan	http://bio2rdf.org/drugbank:DB01030	DDI between Filgrastim and Topotecan - Filgrastim may enhance the adverse/toxic effect of Topotecan.

Tabla 4. Resultados de la consulta SPARQL del ejemplo de la sección 3.3.3 en una tabla, para su mejor comprensión.

3.3.4. Transformación del contenido de aumento en un extracto de HCE normalizado

Esta sección se centra en abordar el paso 6 de la figura 9, que abarca dos tareas: (i) modelar el arquetipo de aumento, que debe contener al menos un elemento para cada elemento de datos del contenido de aumento (es decir, para cada variable en la cláusula SELECT de la consulta SPARQL en la sección 3.3.3), y (ii) definir las especificaciones de mapeo entre el esquema del W3C para resultados de consultas SPARQL y el arquetipo de aumento (es decir, la transformación de datos 1 a nivel de esquema de la figura 8). Una opción para modelar el arquetipo de aumento automáticamente, es programar la generación de un arquetipo cuya raíz sea un CLUSTER, e iterativamente defina un elemento en el CLUSTER por cada variable incluida en el contenido de aumento. De forma similar, la especificación del mapeo es automatizable, estableciendo una relación 1:1 entre las variables de la consulta SPARQL y los elementos del arquetipo de aumento. Sin embargo, para garantizar que los conceptos implicados conservan su significado original, es aconsejable realizar estos pasos manualmente.

La plataforma LinkEHR facilita la especificación de mapeos entre el esquema W3C para resultados de consultas SPARQL y el arquetipo de aumento. LinkEHR incluye un conjunto de herramientas para modelar manualmente los arquetipos de HCE. Las capacidades de mapeo de LinkEHR permiten definir mapeos declarativos de alto nivel que establecen las relaciones entre esquemas de datos. El lenguaje de mapeo especifica cómo calcular un valor para un atributo atómico del esquema de destino utilizando un conjunto de elementos atómicos del esquema de origen. Estas correspondencias de valores se definen mediante un conjunto de

pares, cada par compuesto por una función de transformación y un filtro. El tipo más sencillo de función de transformación es la función de identidad, que copia un valor de origen en un atributo atómico de destino. En los casos en que es necesario transformar los valores atómicos de origen, LinkEHR admite una amplia gama de funciones de transformación. Una vez definidas las especificaciones de mapeo, LinkEHR las implementa un script XQuery capaz de transformar instancias del esquema de origen en instancias del esquema de destino. LinkEHR puede utilizarse para modelar el arquetipo de aumento y para definir las especificaciones de mapeo entre el esquema del W3C para resultados de consultas SPARQL y el arquetipo de aumento. A partir de las especificaciones de mapeo, LinkEHR implementa un script XQuery, cuya ejecución transforma el contenido de aumento en un extracto de HCE normalizado de acuerdo con el arquetipo de aumento (es decir, transformación de datos 1 a nivel de datos en la figura 8).

Ejemplo: El arquetipo de aumento para las interacciones entre medicamentos se modela como un CLUSTER que contiene tres elementos: la descripción de la interacción entre medicamentos, el nombre y el código del medicamento que interactúa (un elemento para cada variable del contenido de aumento ?interactionDescription, ?targetDrugName y ?targetDrugCode, respectivamente). El CLUSTER puede repetirse tantas veces como interacciones farmacológicas se encuentren para el medicamento activo en el extracto inicial de HCE. A continuación, se definen las especificaciones de mapeo mediante funciones de identidad entre los elementos del esquema W3C para resultados de consultas SPARQL y los elementos del arquetipo de aumento. Tanto el modelado del arquetipo de aumento como la definición de las especificaciones de mapeo se realizan utilizando la plataforma LinkEHR. La figura 13 muestra el arquetipo de aumento CEN-EN13606-CLUSTER.DrugDrugInteraction.v1 definido para este ejemplo, y el contenido de aumento resultante del ejemplo de la sección 3.3.3 formateado según este arquetipo.

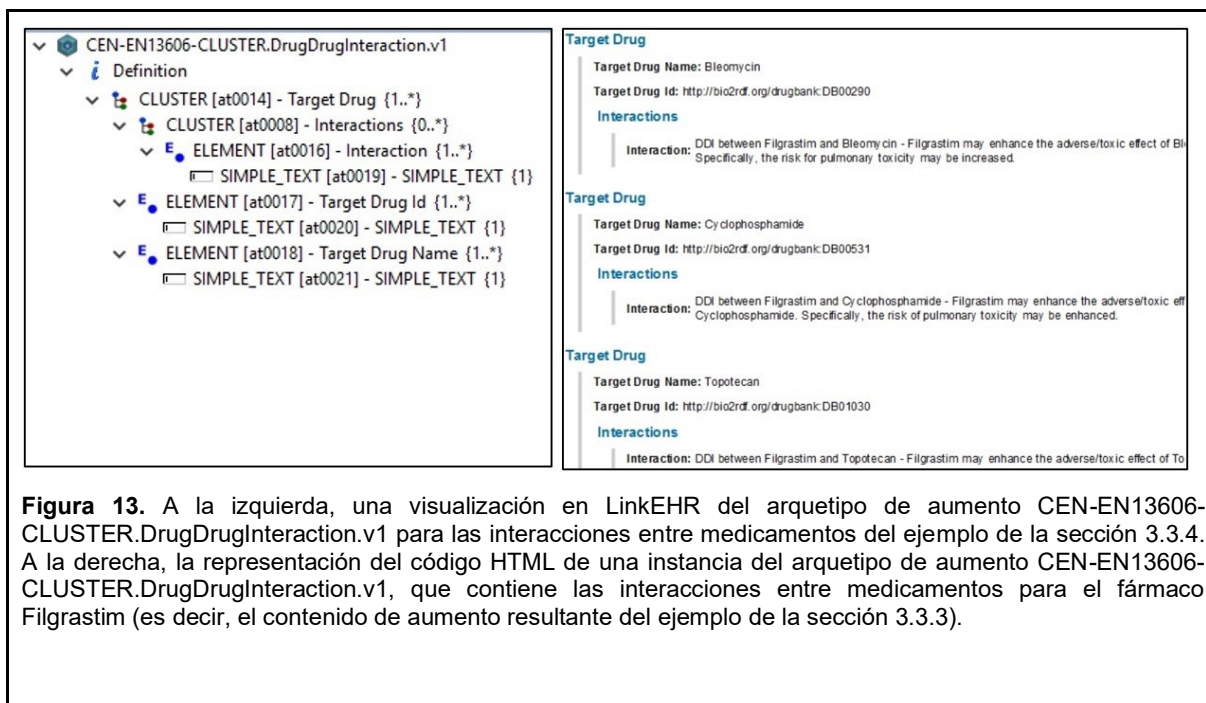


Figura 13. A la izquierda, una visualización en LinkEHR del arquetipo de aumento CEN-EN13606-CLUSTER.DrugDrugInteraction.v1 para las interacciones entre medicamentos del ejemplo de la sección 3.3.4. A la derecha, la representación del código HTML de una instancia del arquetipo de aumento CEN-EN13606-CLUSTER.DrugDrugInteraction.v1, que contiene las interacciones entre medicamentos para el fármaco Filgrastim (es decir, el contenido de aumento resultante del ejemplo de la sección 3.3.3).

3.3.5. Integración del contenido de aumento normalizado y el extracto inicial de la HCE

El arquetipo aumentado comprende el arquetipo inicial y el arquetipo de aumento (paso 7 de la figura 9). Una forma de construir el arquetipo aumentado es especializar el arquetipo inicial incluyendo un SLOT para embeber el arquetipo aumentado. La ruta de acceso al SLOT debe seleccionarse manualmente para garantizar que el arquetipo resultante conserve el significado original. Sin embargo, una vez elegida la ubicación del SLOT, el arquetipo aumentado se modela automáticamente sobre la base del arquetipo inicial, la ruta del SLOT y el arquetipo de aumento (cuyo componente raíz, típicamente un CLUSTER, establece el tipo de dato del SLOT).

Para ayudar en el proceso de elección de la ubicación adecuada del SLOT para el arquetipo de aumento, se presenta un algoritmo que permite identificar qué componentes del arquetipo inicial son válidos para contener el SLOT, de acuerdo con las restricciones del modelo de referencia y con la cardinalidad de los componentes del arquetipo inicial. El algoritmo consiste en identificar el componente más cercano a la EDI, dentro del arquetipo inicial, tal que pueda albergar elementos del mismo tipo que el componente raíz del arquetipo de aumento (considerando las restricciones del modelo de referencia y del modelo de arquetipos). El resultado de este algoritmo es un conjunto de rutas candidatas ordenadas por proximidad a las EDI del arquetipo inicial. A continuación se describe este algoritmo en pseudo código:

Variables de entrada:

- *Root_Type*: tipo de componente del MR ISO13606 que corresponde con el elemento raíz del arquetipo de aumento.
- *EHR_Archetype_Elements*: conjunto de elementos del arquetipo inicial
- *EOI_Paths*: rutas a las EDIs en el arquetipo inicial
- *Preferred_Slot_Path*: ruta a la localización del slot para el arquetipo de aumento en el arquetipo inicial, proporcionada por el usuario (opcional, puede estar vacía).

Variable de salida:

- *Slot_Locations*: rutas candidatas para referenciar al arquetipo de aumento en el arquetipo aumentado.

Inicialización:

- *Slot_Paths*: conjunto de rutas candidatas para la localización del slot en el arquetipo aumentado (inicialmente vacío).

Proceso:

para cada element **en** *EHR_Archetype_Elements* **hacer:**

si element.type puede contener un elemento de tipo *Root_Type*

y element.max_cardinality > element.cardinality **entonces:**

Slot_Paths.add(element.path)

fin si

fin para

si *Slot_Paths* contiene *Preferred_Slot_Path* **entonces:**

Slot_Locations = Preferred_Slot_Path

si no

Slot_Locations = ordenar_por_proximidad(Slot_Paths, EOI_Paths)

fin si

devolver *Slot_Locations*

El valor devuelto *Slot_Locations* proporciona un conjunto de rutas del arquetipo inicial que son candidatas válidas para albergar el arquetipo de aumento. La ruta final del SLOT para el arquetipo de aumento debe seleccionarse manualmente del conjunto de rutas devueltas en *Slot_Locations*. A continuación, el arquetipo aumentado puede modelarse como una

especialización del arquetipo inicial, que contiene los mismos elementos que el arquetipo inicial más un SLOT adicional para el arquetipo de aumento, ubicado en la ruta seleccionada del conjunto Slot_Locations.

La plataforma LinkEHR proporciona herramientas para modelar el arquetipo aumentado y para definir las reglas de mapeo entre el arquetipo inicial, el arquetipo de aumento y el arquetipo aumentado (es decir, la transformación de datos 2 a nivel de esquema en la figura 8). En este caso, las especificaciones de mapeo se basan en funciones de identidad e inclusiones de SLOT. Por último, la plataforma LinkEHR sirve para transpilar las especificaciones de mapeo en un script XQuery, cuya ejecución transforma instancias del arquetipo inicial, junto con el contenido de aumento normalizado, en un extracto de HCE aumentado, expresado como una instancia del arquetipo aumentado (es decir, la transformación de datos 2 a nivel de datos en la figura 8).

Obsérvese que un arquetipo aumentado podría modelarse para albergar varios arquetipos de aumento incluyendo SLOT adicionales para ellos en la etapa de modelado. En tales casos, sería necesario repetir los pasos de la sección 3.3.5 para cada arquetipo de aumento que se incluya en el arquetipo aumentado.

Ejemplo: El último paso del ejemplo es generar el arquetipo aumentado CEN-EN13606-ENTRY.MedicacionAumentada.v1, como una especialización del arquetipo inicial CEN-EN13606-ENTRY.Medicacion.v1. En este ejemplo, el arquetipo aumentado contiene el arquetipo de aumento CEN-EN13606-CLUSTER.Drug-DrugInteraction.v1 y el arquetipo inicial.

Para identificar la ruta del SLOT para el arquetipo de aumento, se ejecuta el algoritmo introducido en la sección 3.3.5 utilizando las entradas:

Root_Type: CLUSTER

EHR_Archetype_Elements: Elementos en CEN-EN13606-ENTRY.Medicacion.v1

EOI_Paths: Ruta de acceso al valor codificado de la medicación (EDI) en el arquetipo inicial
/articulos[archetype_id="at0005"]/partes[archetype_id="at0010"]/valor

Preferred_Slot_Path: /items[archetype_id="at0005"]

Entre las rutas candidatas devueltas por el algoritmo, la primera es la ruta /items[archetype_id="at0005"]. Esta ruta es válida y adecuada para incluir un SLOT en una especialización del arquetipo inicial. La ruta es válida porque apunta a un componente de tipo CLUSTER, que puede contener el arquetipo de aumento (cuyo componente raíz es también un CLUSTER) según el modelo de referencia ISO13606. La ruta es adecuada

porque apunta al componente contenedor del valor codificado de la medicación, que es la EDI utilizada para obtener el contenido de aumento. Ubicar el contenido de aumento cerca de los valores codificados de EDI (la medicación) enriquece los datos de la medicación con información sobre las interacciones entre medicamentos.

Por último, el arquetipo aumentado puede generarse automáticamente a partir del arquetipo inicial, el arquetipo de aumento y la ruta del SLOT para el arquetipo de aumento. La figura 14 muestra el arquetipo aumentado resultante en contraste con el arquetipo inicial.

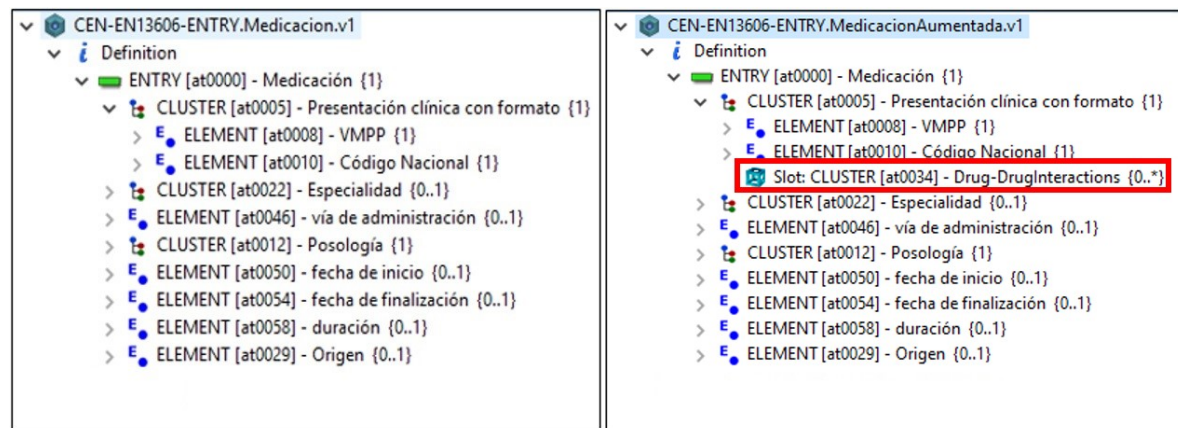


Figura 14. Visualización de arquetipos en LinKEHR. A la izquierda, el arquetipo inicial CEN-EN13606-ENTRY.Medicacion.v1. A la derecha, el arquetipo aumentado CEN-EN13606-ENTRY.MedicacionAumentada.v1.

3.4. Resultados

Definir y modelar una HCE aumentada es un proceso de diseño que conlleva la realización manual de algunas tareas de modelado. Sin embargo, una vez que se define una HCE aumentada para un determinado arquetipo inicial, puede aplicarse a cualquier instancia de dicho arquetipo. Esta sección presenta el diseño e implementación de un software para ejecutar las transformaciones de datos y las consultas necesarias para generar extractos de HCE aumentada a partir de extractos iniciales de HCE, una vez realizadas las tareas de modelado necesarias. El flujo de trabajo (figura 15) parte del extracto inicial de HCE y del identificador de la HCE aumentada, y finaliza devolviendo un extracto de HCE aumentada.

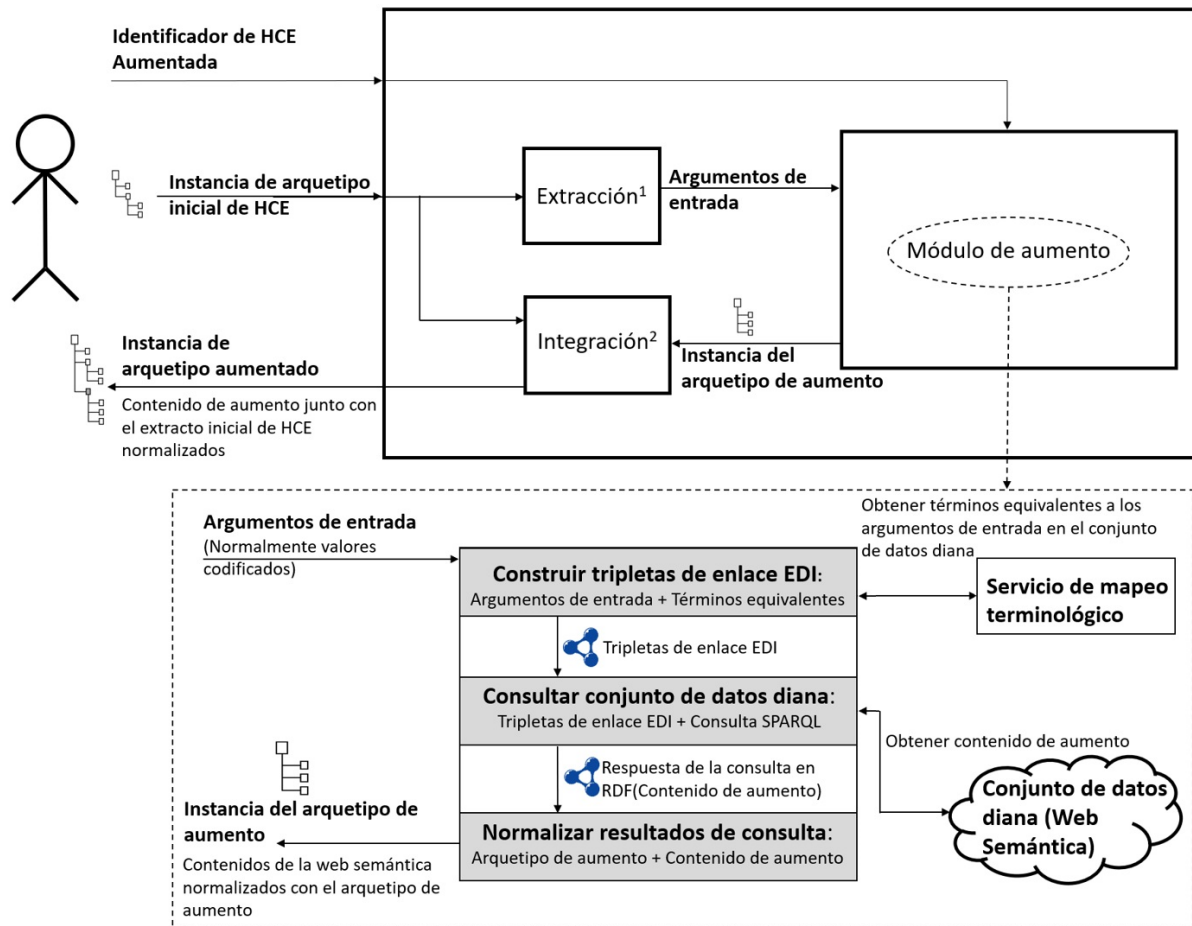


Figura 15. Descripción del proceso para obtener un extracto de HCE aumentada (es decir, el ciclo de ejecución de la HCE aumentada). 1. Extracción: Lee valores codificados EDI y los proporciona al módulo de aumento. 2. Integración: Genera instancias del arquetipo aumentado combinando la instancia del arquetipo de HCE y la instancia del arquetipo de aumento.

El módulo de extracción identifica y obtiene los valores codificados EDI del extracto inicial de HCE. El módulo de aumento toma los valores EDI como entradas y ejecuta el procedimiento descrito en la sección 3.3.2 para construir las tripletas de enlace EDI. Estas tripletas se inyectan en el grafo de una consulta SPARQL para obtener el contenido de aumento de un conjunto de datos LOD, como se explica en la sección 3.3.3. El contenido de aumento se normaliza de acuerdo con el arquetipo de aumento (sección 3.3.4). Por último, el módulo de integración genera el extracto de HCE aumentada a partir del extracto inicial de HCE y el contenido de aumento normalizado, como se describe en la sección 3.3.5.

Se ha implementado un caso de uso utilizando el lenguaje de scripting XQuery para desarrollar el módulo de Extracción, el módulo de Integración y los pasos específicos realizados por el módulo de aumento: construcción de tripletas de enlace EDI, y normalización del contenido de aumento. Las ventajas de utilizar XQuery son, en primer lugar, el aprovechamiento de los scripts de transformación que la plataforma LinkEHR genera al transpilar las especificaciones de mapeo (es decir, la transformación de datos 1 y 2 a nivel de datos en la figura 8). En segundo lugar, el hecho de que XQuery esté diseñado para la

extracción y manipulación de datos XML facilita las tareas de extracción de valores EDI en el módulo de Extracción y la construcción de tripletas de enlace EDI en el módulo de Aumento. Además, el lenguaje de consulta SPARQL se emplea en el módulo de aumento para obtener el contenido de aumento, y el lenguaje de programación Java se utiliza en el caso de uso para orquestar el flujo de trabajo y las comunicaciones entre los módulos implicados. El desarrollo cuenta con un repositorio de aumentos de HCE disponibles, que se ha omitido en la figura 15 en aras de la claridad. Para cada aumento de HCE, este repositorio contiene:

- I. El script XQuery para construir las tripletas de enlace EDI a partir de los valores codificados EDI
- II. La consulta SPARQL para obtener el contenido del aumento
- III. El arquetipo de aumento
- IV. El script XQuery para normalizar el contenido del aumento según el arquetipo de aumento
- V. Una lista de arquetipos de HCE iniciales que se pueden enriquecer con este aumento de HCE. Para cada arquetipo inicial, el repositorio contiene:
 - A. Rutas del arquetipo que apuntan a las EDI en el arquetipo inicial, necesarias para extraer los valores codificados EDI de los extractos iniciales de HCE
 - B. El arquetipo aumentado
 - C. El script XQuery para generar extractos de HCE aumentados a partir de extractos iniciales de HCE y contenidos de aumento normalizados

Para configurar un aumento de HCE, es necesario añadir manualmente esta información (I - V) al repositorio. Al tratarse mayoritariamente de datos XML, esquemas XML, y scripts XQuery, la tecnología escogida para el repositorio ha sido eXistDB, una base de datos documental basada en XML.

El caso de uso está pensado para facilitar la integración con sistemas de HCE que cumplan con estándares basados en el modelo dual de HCE. La integración debe realizarse por profesionales de la informática médica. Una vez integrado, permite a los usuarios (por ejemplo, médicos u otros profesionales de la salud) acceder automáticamente a los extractos de HCE aumentada de la misma forma que acceden a la HCE de los pacientes. El sistema comprueba el repositorio en busca de aumentos de HCE adecuados, mediante la búsqueda de arquetipos de HCE iniciales en (V) que coincidan con el extracto de HCE inicial. A continuación, el proceso se inicia como se describe en la figura 15 para los aumentos de HCE identificados. Los resultados comprenden contenidos de aumento formateados como extractos de HCE, que se envían al sistema de HCE para ser representados en un visor de HCE normalizada, proporcionando a los usuarios extractos de HCE aumentados con contenidos de aumento específicos bajo demanda.

3.5. Caso de uso

El arquetipo de HCE resumida, definido por el Ministerio de Sanidad de España (CEN-EN13606-COMPOSITION.HistoriaClinicaResumida.v1) para la comunicación de datos sanitarios entre instituciones sanitarias regionales [118], es un arquetipo inicial adecuado para ilustrar la idea de la HCE aumentada. Abarca elementos relativos a las HCE de los pacientes, tales como problemas actuales y tratamientos en curso de los pacientes, susceptibles de ser utilizados como EDIs para enriquecer el arquetipo de HCE resumida con varios aumentos de HCE. Esta sección muestra el proceso de diseño del arquetipo de HCE resumida aumentada (CEN-EN13606-COMPOSITION.HistoriaClinicaResumidaAumentada.v1) desde cero, como resultado de enriquecer el arquetipo de HCE resumida con aumentos de HCE relacionados con las interacciones entre medicamentos y los posibles tratamientos de las enfermedades del paciente. Las EDI seleccionadas del arquetipo de HCE resumida son la medicación activa y los problemas o episodios activos de los pacientes, situados en las rutas

```
/content[archetype_id="at0043"]/members[archetype_id="at0024"]/members[archetype_id="at0053"]/members[archetype_id="at0054"]/items[archetype_id="at0172"]/parts[archetype_id="at0181"]/value
```

y

```
/content[archetype_id="at0043"]/members[archetype_id="at0042"]/members[archetype_id="at0046"]/items[archetype_id="at0158"]/value,
```

respectivamente. Estas EDI contienen valores codificados de la terminología SNOMED-CT, susceptibles de ser aumentados con información adicional potencialmente interesante, procedente de fuentes de información heterogéneas en formato de web semántica. La figura 16 muestra el arquetipo inicial y un extracto inicial de HCE con las EDI seleccionadas.

La información de medicamentos para enriquecer la sección de medicación activa del paciente con interacciones entre medicamentos se encuentra en la base de datos DrugBank. Esta información puede resultar útil para el especialista a la hora de prescribir fármacos adicionales a un paciente con medicación activa. La información sobre tratamientos de enfermedades específicas y sus efectos, para enriquecer la sección de problemas o episodios activos del paciente, está disponible en la ontología NDFRT. Esta información puede resultar útil para el especialista tras diagnosticar los problemas del paciente. Los EDE entre los términos codificados de las terminologías de origen (SNOMED-CT) y destino (DrugBank, NDFRT) se identifican utilizando el servicio de mapeo terminológico de BioPortal. Las tripletas de enlace EDI para establecer mapeos conceptuales entre las terminologías de origen y destino se construyen consultando a dicho servicio por los términos EDI. Se requiere una representación semántica de las terminologías de destino para obtener el contenido de aumento a través de consultas SPARQL. La LODC contiene una representación semántica

de la base de datos DrugBank (Bio2RDF DrugBank), mientras que la terminología NDFRT expresada como LOD se encuentra en el repositorio ontológico BioPortal.

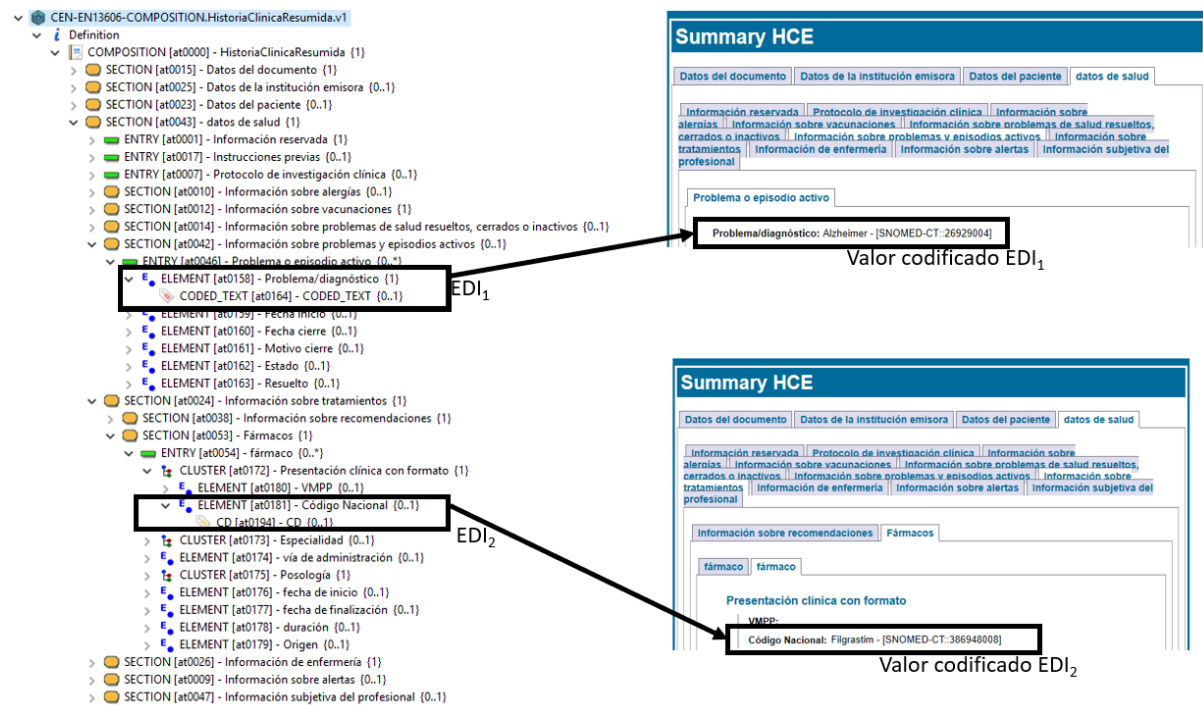


Figura 16. A la izquierda, una visualización mediante LinkEHR del arquetipo inicial CEN-EN13606-COMPOSITION.HistoriaClinicaResumida.v1. A la derecha, dos vistas de la representación del código HTML del extracto inicial de HCE. Tanto en el arquetipo inicial como en el extracto de HCE inicial, se marcan los problemas o episodios activos del paciente (EDI₁) y la medicación activa del paciente (EDI₂), así como sus respectivos valores codificados.

El procedimiento automático descrito en la sección 3.3.2 sirve para construir las triplas de enlace EDI tomando varios argumentos como entradas, de las cuales solo es necesario proporcionar el extracto inicial de HCE en tiempo de ejecución. Los demás argumentos se definen como parte del aumento de HCE en tiempo de diseño. Las tablas 5 y 6 enumeran los argumentos necesarios para ambos aumentos. En la tabla 5, nótese que el prefijo de la URL de los EDE (por ejemplo, <http://bio2rdf.org/drugbank:>) difiere de la URL de la fuente externa devuelta por el servicio de mapeo (por ejemplo, <http://www.drugbank.ca/drugs>). En este caso, el prefijo de la URL de los EDE en el conjunto de datos de destino puede identificarse consultando descriptores VOID, como se explica en la sección 3.3.1.

	Variable	Valor
Entradas	Extracto inicial de la HCE	Una instancia del arquetipo de HCE resumida proporcionada en tiempo de ejecución
	Ruta de acceso a la EDI en el arquetipo inicial	/content[archetype_id="at0043"] /members[archetype_id="at0024"] /members[archetype_id="at0053"] /members[archetype_id="at0054"] /items[archetype_id="at0172"] /parts[archetype_id="at0181"]/value
	La URL base del predicado para la tripleta de enlace EDI	http://www.w3.org/2004/02/skos/core#exactMatch
	El nombre de la terminología de origen empleada por los valores codificados EDI	SNOMEDCT
	El identificador URL de la fuente de datos externa devuelto por el servicio de mapeo terminológico	http://www.drugbank.ca/drugs
	El prefijo URL de los EDE en el conjunto de datos de destino (representación semántica de la fuente externa)	http://bio2rdf.org/drugbank:
Salidas	Las tripletas de enlace EDI	Tripletas RDF que relacionan las EDI del extracto inicial de HCE con EDE del conjunto de datos de destino

Tabla 5. Entradas y salidas del procedimiento automático para construir las tripletas de enlace EDI para el aumento de HCE de las interacciones medicamentosas.

Las tripletas de enlace EDI resultantes para el aumento de HCE de interacciones medicamentosas se generan utilizando las entradas de la tabla 5 y el extracto inicial de HCE introducido en la figura 16, cuyo valor codificado EDI (medicación activa del paciente) es Filgrastim-[SNOMED-CT::386948008]. Las tripletas de enlace EDI para el aumento de HCE

de tratamientos de problemas o episodios activos del paciente se construyen con las entradas de la tabla 6 y el extracto inicial de la HCE introducido en la figura 16, cuyo valor codificado EDI (episodios activos del paciente) es Alzheimer-[SNOMED-CT::26929004]. La tabla 7 resume la información relacionada con las tripletas de enlace EDI obtenidas para este caso de uso.

	Variable	Valor
Entradas	Extracto inicial de la HCE	Una instancia del arquetipo de HCE resumida proporcionada en tiempo de ejecución
	Ruta de acceso a la EDI en el arquetipo inicial	/content[archetype_id="at0043"] /members[archetype_id="at0042"] /members[archetype_id="at0046"] /artículos[archetype_id="at0158"]/valor
	La URL base del predicado para la tripleta de enlace EDI	http://www.w3.org/2004/02/skos/core#exactMatch
	El nombre de la terminología de origen empleada por los valores codificados EDI	SNOMEDCT
	El identificador URL de la fuente externa devuelto por el servicio de mapeo terminológico	http://purl.bioontology.org/ontology/ND FRT
	El prefijo URL de los EDE en el conjunto de datos de destino (representación semántica de la fuente externa)	http://purl.bioontology.org/ontology/ND FRT
Salidas	Las tripletas de enlace EDI	Tripletas RDF que relacionan las EDI del extracto inicial de HCE con EDE del conjunto de datos de destino

Tabla 6. Entradas y salidas del procedimiento automático para construir las tripletas de enlace EDI para el aumento de HCE de tratamientos relacionados con episodios activos del paciente.

Aumento de HCE	EDI	Valor codificado EDI	Tripleta de enlace EDI
Interacciones entre fármacos	Medicación activa del paciente	Filgrastim- [SNOMED-CT::386948008]	<p><http://purl.bioontology.org/ontology/SNOMEDCT/386948008></p> <p><http://www.w3.org/2004/02/skos/core#exactMatch></p> <p><http://bio2rdf.org/drugbank:DB00099></p> <p>En este caso, DB00099 es el código terminológico en DrugBank equivalente al código 386948008 de SNOMED-CT. Ambos representan el mismo concepto (el fármaco Filgrastim). La equivalencia entre estos códigos es proporcionada por el servicio de mapeos terminológico.</p>
Tratamientos para los problemas o episodios activos del paciente	Problemas o episodios activos del paciente	Alzheimer- [SNOMED-CT::26929004]	<p><http://purl.bioontology.org/ontology/SNOMEDCT/26929004></p> <p><http://www.w3.org/2004/02/skos/core#exactMatch></p> <p><http://purl.bioontology.org/ontology/NDFRT/N0000000363></p> <p>En este caso, N0000000363 es el código terminológico en NDFRT equivalente al código 26929004 de SNOMED-CT. Ambos representan el mismo concepto (enfermedad de Alzheimer). La equivalencia entre estos códigos es proporcionada por el servicio de mapeos terminológico.</p>

Tabla 7. Tripletas de enlace EDI generadas para cada aumento de HCE, a partir del extracto inicial de HCE introducido en la figura 16 y los argumentos descritos en la tabla 5 y la tabla 6, respectivamente.

Las tripletas de enlace EDI se incorporan en tiempo de ejecución a una consulta SPARQL como parte del grafo de consulta para obtener el contenido de aumento. La figura 17 muestra

las consultas SPARQL para obtener las interacciones entre medicamentos y tratamientos para los problemas o episodios activos del paciente de los aumentos de HCE.

<pre> PREFIX sct: <http://purl.bioontology.org/ontology/SNOMEDCT/> PREFIX sch: <http://schema.org/> PREFIX db: <http://bio2rdf.org/drugbank:> PREFIX dbv: <http://bio2rdf.org/drugbank_vocabulary:> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX dcterms: <http://purl.org/dc/terms/> PREFIX skos: <http://www.w3.org/2004/02/skos/core#> SELECT DISTINCT ?reference_drug_name ?reference_drug_code_sct ?reference_drug_code_db ?target_drug_name ?target_drug_code_db ?interaction_description ?interaction_code_db WHERE { ?reference_drug_code_sct skos:exactMatch ?reference_drug_code_db . SERVICE <http://bio2rdf.org/sparql> { ?interaction_code_db rdf:type dbv:Drug-Drug-Interaction . ?reference_drug_code_db dbv:ddi-interactor-in ?interaction_code_db . ?target_drug_code_db dbv:ddi-interactor-in ?interaction_code_db . ?target_drug_code_db dcterms:title ?target_drug_name . ?reference_drug_code_db dcterms:title ?reference_drug_name . ?interaction_code_db dcterms:title ?interaction_description . FILTER(?reference_drug_code_db != ?target_drug_code_db) } } </pre>	<pre> PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> PREFIX ndfirt: <http://purl.bioontology.org/ontology/NDFIRT/> PREFIX skos: <http://www.w3.org/2004/02/skos/core#> PREFIX sct: <http://purl.bioontology.org/ontology/SNOMEDCT/> PREFIX skos: <http://www.w3.org/2004/02/skos/core#> SELECT DISTINCT ?disease_code_sct ?disease_code_ndfirt ?disease_description ?treatment_code_ndfirt ?treatment_description ?effect_code_ndfirt ?effect_description WHERE { ?disease_code_sct skos:exactMatch ?disease_code_ndfirt . SERVICE <http://sparql.bioontology.org/sparql?apikey=[APIKEY]> { ?disease_code_ndfirt ndfirt:may_be_treated_by ?treatment_code_ndfirt . ?treatment_code_ndfirt ndfirt:has_product_component ?prod . ?treatment_code_ndfirt skos:prefLabel ?treatment_description . ?treatment_code_ndfirt ndfirt:has_physiologic_effect ?effect_code_ndfirt . ?effect_code_ndfirt skos:prefLabel ?effect_description . ?disease_code_ndfirt skos:prefLabel ?disease_description . } } </pre>
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Figura 17. Consultas SPARQL para obtener los contenidos de aumento. A la izquierda, la consulta para obtener interacciones entre fármacos. A la derecha, la consulta para obtener tratamientos para los problemas o episodios activos del paciente.

Los arquetipos de aumento para normalizar el contenido de aumento se modelan en la etapa de diseño. La figura 18 muestra el arquetipo de aumento de las interacciones entre fármacos (CEN-EN13606-CLUSTER.DrugDrugInteraction.v1) y el arquetipo de aumento de tratamientos de enfermedades (problemas o episodios activos) del paciente (CEN-EN13606-CLUSTER.DiseaseTreatment.v1). Estos arquetipos de aumento se han modelado utilizando LinkEHR. En ambos arquetipos de aumento, el elemento raíz es un CLUSTER que contiene al menos un elemento por cada variable en la consulta SPARQL para albergar el contenido del aumento.

La figura 19 muestra el arquetipo aumentado CEN-EN13606-COMPOSITION.HistoriaClinicaResumidaAumentada.v1, que resulta de combinar los arquetipos de aumento y el arquetipo inicial. Se ha utilizado LinkEHR tanto para modelar los arquetipos requeridos como para definir las relaciones de mapeo entre estos arquetipos y sus correspondientes fuentes de información. La tabla 8 resume los esquemas de datos involucrados en los mapeos de identidad necesarios para generar las instancias de la HCE resumida aumentada. Nótese que el único arquetipo que existe previamente a la definición de una HCE aumentada es el arquetipo inicial (CEN-EN13606-COMPOSITION.HistoriaClinicaResumida.v1). Los arquetipos de aumento y el arquetipo aumentado se modelan como parte del proceso de diseño de la HCE aumentada.

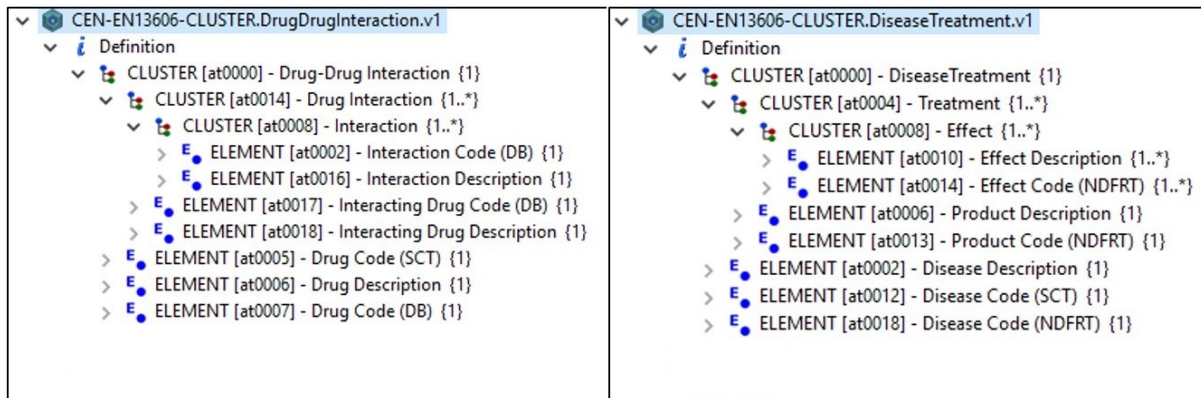


Figura 18. Representación de arquetipos de aumento en LinkEHR. A la izquierda, el arquetipo de aumento CEN-EN13606-CLUSTER.DrugDrugInteraction.v1 para el aumento de HCE de las interacciones entre fármacos. A la derecha, el arquetipo de aumento CEN-EN13606-CLUSTER.DiseaseTreatment.v1 para el aumento de HCE de los tratamientos relacionados con las problemas o episodios activos del paciente.



Figura 19. Representación de arquetipos en LinkEHR. A la izquierda, el arquetipo inicial CEN-EN13606-COMPOSITION.HistoriaClinicaResumida.v1. Se resaltan las EDI utilizadas en este caso de uso. A la derecha, el arquetipo aumentado CEN-EN13606-COM-POSITION.HistoriaClinicaResumidaAumentada.v1, similar al arquetipo inicial pero incluyendo dos SLOT para los arquetipos de aumento.

La etapa de diseño concluye una vez que se han identificado y definido los recursos necesarios: Rutas de EDI en el arquetipo inicial donde está ubicado el contenido a aumentar, conjunto de datos de destino, servicio de mapeo terminológico, consultas SPARQL, argumentos para construir las tripletas de enlace EDI, arquetipos de aumento, arquetipo aumentado y sus correspondientes especificaciones de mapeo. En tiempo de ejecución, solo se requiere una instancia del arquetipo inicial para iniciar el proceso y generar una instancia del arquetipo aumentado. La figura 20 muestra los contenidos de aumento para el caso de uso, normalizados según los arquetipos de aumento, y la figura 21 ilustra la instancia del arquetipo aumentado generada para el caso de uso.

Recursos	Esquema(s) de origen	Esquema de destino
Contenido de aumento de interacciones entre fármacos	Esquema XML para los resultados de la consulta SPARQL	CEN-EN13606-CLUSTER.DrugDrugInteraction.v1
Contenido de aumento de tratamientos de los episodios activos del paciente	Esquema XML para los resultados de la consulta SPARQL	CEN-EN13606-CLUSTER.DiseaseTreatment.v1
Resumen de HCE aumentado	<ul style="list-style-type: none"> - El arquetipo de HCE resumida CEN-EN13606-COMPOSITION.HistoriaClinicaResumida.v1 - El arquetipo de interacciones medicamentosas CEN-EN13606-CLUSTER.DrugDrugInteraction.v1 - El arquetipo de tratamientos de enfermedad CEN-EN13606-CLUSTER.DiseaseTreatment.v1 	CEN-EN13606-COMPOSITION.HistoriaClinicaResumidaAumentada.v1

Tabla 8. Esquemas de datos involucrados en los mapeos para generar instancias de la HCE resumida aumentada.

<p>Drug-Drug Interaction</p> <p>Drug Code (SCT): http://purl.bioontology.org/ontology/SNOMEDCT/386948008</p> <p>Drug Description: Filgrastim</p> <p>Drug Code (DB): http://bio2rdf.org/drugbank:DB00099</p> <p>Drug Interaction</p> <p>Interacting Drug Description: Bleomycin</p> <p>Interacting Drug Code (DB): http://bio2rdf.org/drugbank:DB00290</p> <p>Interaction</p> <p>Interaction Description: DDI between Filgrastim and Bleomycin - Filgrastim may enhance the adverse/toxic effect of Bleomycin. Specifically, the risk for pulmonary toxicity may be increased.</p> <p>Interaction Code (DB): http://bio2rdf.org/drugbank_resource:DB00099_DB00290</p> <p>Drug Interaction</p> <p>Interacting Drug Description: Cyclophosphamide</p> <p>Interacting Drug Code (DB): http://bio2rdf.org/drugbank:DB00531</p> <p>Interaction</p> <p>Interaction Description: DDI between Filgrastim and Cyclophosphamide - Filgrastim may enhance the adverse/toxic effect of Cyclophosphamide. Specifically, the risk of pulmonary toxicity may be enhanced.</p> <p>Interaction Code (DB): http://bio2rdf.org/drugbank_resource:DB00099_DB00531</p> <p>Drug Interaction</p> <p>Interacting Drug Description: Topotecan</p> <p>Interacting Drug Code (DB): http://bio2rdf.org/drugbank:DB01030</p> <p>Interaction</p> <p>Interaction Description: DDI between Filgrastim and Topotecan - Filgrastim may enhance the adverse/toxic effect of Topotecan.</p> <p>Interaction Code (DB): http://bio2rdf.org/drugbank_resource:DB00099_DB01030</p>	<p>Disease Treatment</p> <p>Disease Description: Alzheimer Disease [Disease/Finding]</p> <p>Disease Code (SCT): http://purl.bioontology.org/ontology/SNOMEDCT/26929004</p> <p>Disease Code (NDFRT): http://purl.bioontology.org/ontology/NDFRT/N0000000363</p> <p>Treatment</p> <p>Product Description: RIVASTIGMINE 9.5MG/24HR PATCH [VA Product]</p> <p>Product Code (NDFRT): http://purl.bioontology.org/ontology/NDFRT/N0000177418</p> <p>Effect</p> <p>Effect Description: Increased Central Nervous System Acetylcholine Activity [PE]</p> <p>Effect Code (NDFRT): http://purl.bioontology.org/ontology/NDFRT/N00000009178</p> <p>Treatment</p> <p>Product Description: DIVALPROEX NA 125MG TAB,EC [VA Product]</p> <p>Product Code (NDFRT): http://purl.bioontology.org/ontology/NDFRT/N0000154279</p> <p>Effect</p> <p>Effect Description: Increased GABA Activity [PE]</p> <p>Effect Code (NDFRT): http://purl.bioontology.org/ontology/NDFRT/N00000009325</p> <p>Treatment</p> <p>Product Description: VALPROIC ACID 250MG CAP,SA [VA Product]</p> <p>Product Code (NDFRT): http://purl.bioontology.org/ontology/NDFRT/N0000178119</p> <p>Effect</p> <p>Effect Description: Increased GABA Activity [PE]</p> <p>Effect Code (NDFRT): http://purl.bioontology.org/ontology/NDFRT/N00000009325</p>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Figura 20. Representación HTML del contenido del aumento. A la izquierda, el contenido de aumento de HCE de interacciones entre fármacos, formateado según el arquetipo de aumento CEN-EN13606-CLUSTER.DrugDrugInteraction.v1. A la derecha, el contenido de aumento de HCE de tratamientos de problemas o episodios activos del paciente, formateado según el arquetipo de aumento CEN-EN13606-CLUSTER.DiseaseTreatment.v1.

<p>Summary HCE</p> <p>Datos del documento Datos de la institución emisora Datos del paciente datos de salud</p> <p>Información reservada Protocolo de investigación clínica Información sobre alergias Información sobre vacunaciones Información sobre problemas de salud resueltos, cerrados o inactivos Información sobre problemas v episodios activos Información sobre tratamientos Información de enfermería Información sobre alertas Información subjetiva del profesional</p> <p>Información sobre recomendaciones Fármacos</p> <p>fármaco fármaco</p> <p>Presentación clínica con formato</p> <p>VMPP: Código Nacional: Filgrastim - [SNOMED-CT: 386948008]</p> <p>[...]</p> <p>Drug-Drug Interaction</p> <p>Drug Code (SCT): http://purl.bioontology.org/ontology/SNOMEDCT/386948008</p> <p>Drug Description: Filgrastim</p> <p>Drug Code (DB): http://bio2rdf.org/drugbank:DB00099</p> <p>Drug Interaction</p> <p>Interacting Drug Description: Bleomycin</p> <p>Interacting Drug Code (DB): http://bio2rdf.org/drugbank:DB00290</p> <p>Interaction</p> <p>Interaction Description: DDI between Filgrastim and Bleomycin - Filgrastim may enhance the adverse/toxic effect of Bleomycin. Specifically, the risk for pulmonary toxicity may be increased.</p> <p>Interaction Code (DB): http://bio2rdf.org/drugbank_resource:DB00099_DB00290</p> <p>[...]</p>	<p>Summary HCE</p> <p>Datos del documento Datos de la institución emisora Datos del paciente datos de salud</p> <p>Información reservada Protocolo de investigación clínica Información sobre alergias Información sobre vacunaciones Información sobre problemas de salud resueltos, cerrados o inactivos Información sobre problemas v episodios activos Información sobre tratamientos Información de enfermería Información sobre alertas Información subjetiva del profesional</p> <p>Problema o episodio activo</p> <p>Problema/diagnóstico: Alzheimer - [SNOMED-CT: 26929004]</p> <p>Fecha inicio: 2018-03-30</p> <p>Fecha cierre: 2018-03-30</p> <p>Motivo cierre:</p> <p>Estado:</p> <p>Resultado: Sí</p> <p>[...]</p> <p>Disease Treatment</p> <p>Disease Description: Alzheimer Disease [Disease/Finding]</p> <p>Disease Code (SCT): http://purl.bioontology.org/ontology/SNOMEDCT/26929004</p> <p>Disease Code (NDFRT): http://purl.bioontology.org/ontology/NDFRT/N0000000363</p> <p>Treatment</p> <p>Product Description: RIVASTIGMINE 9.5MG/24HR PATCH [VA Product]</p> <p>Product Code (NDFRT): http://purl.bioontology.org/ontology/NDFRT/N0000177418</p> <p>Effect</p> <p>Effect Description: Increased Central Nervous System Acetylcholine Activity [PE]</p> <p>Effect Code (NDFRT): http://purl.bioontology.org/ontology/NDFRT/N00000009178</p> <p>[...]</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Figura 21. Representación HTML del extracto de HCE resumida aumentada, una instancia del arquetipo aumentado CEN-EN13606-COMPOSITION.HistoriaClinicaResumidaAumentada.v1 que incluye el contenido de aumento de las interacciones entre fármacos (izquierda) y los aumentos de HCE de tratamientos para los episodios activos del paciente (derecha).

3.6. Discusión

Varios trabajos anteriores han abordado la combinación de la HCE con otras fuentes de información [16, 89, 91, 92, 98, 100, 119, 120], la mayoría de los cuales se han centrado en la conversión de los datos de la HCE a formatos de la web semántica. La estrategia tras estos trabajos es razonar con los datos, y así inferir nuevos hechos a partir de la HCE del paciente. Sin embargo, los formatos de la web semántica no están soportados de forma nativa por los sistemas de HCE. Nuestro trabajo introduce un enfoque en la dirección opuesta, con el objetivo de aportar contenidos adicionales de fuentes de la web semántica a la HCE de un determinado paciente, de forma que los sistemas y servicios de información sanitarios normalizados puedan aprovechar estos contenidos.

Otros trabajos abordan métodos para incorporar información adicional potencialmente relevante a la HCE [101–103, 121]. Sin embargo, esta información adicional o contenido de aumento no se normaliza según un estándar de HCE, lo que conduce a un grado limitado de interoperabilidad, en contraste con el enfoque de HCE aumentada descrito en esta tesis, en el que el contenido de fuentes externas se normaliza y se integra como parte de la HCE. Esto permite procesar la HCE aumentada en transformaciones posteriores como una HCE normalizada, gestionando conjuntamente el contenido de aumento y los datos de la HCE del paciente como la misma entidad (es decir, la HCE), lo que puede ser especialmente útil cuando se alimenta la HCE con datos de otros sistemas de información, como sistemas de HCE de terceros, SADC o GCI. La solución Infobuttons [101], incorporada como parte del estándar HL7 (HL7's Context-Aware Knowledge Retrieval), merece mención especial, ya que permite recuperar información externa desde un sistema de HCE que implemente la especificación del estándar HL7. Sin embargo, el trabajo de esta tesis mejora dicho enfoque de dos maneras. En primer lugar, la metodología propuesta permite recuperar información relevante de la web semántica, lo que ofrece ventajas significativas. La especificación de Infobuttons no establece cómo integrar contenidos de la web semántica en la HCE. Además, la solución propuesta define un método para transformar la información de la web semántica en extractos de HCE normalizados, mejorando la interoperabilidad con otros sistemas de HCE estándar basados en el modelo dual. Esto contrasta con el enfoque de Infobuttons, que no introduce ningún método para recuperar los datos adicionales de acuerdo a arquetipos de HCE, lo que limita la interoperabilidad de la información adicional con otros sistemas de HCE basados en el modelo dual. Dado que el método de HCE aumentada no está diseñado para trabajar con HL7, representa una alternativa complementaria a Infobuttons para enriquecer la HCE normalizada de acuerdo con modelos duales.

Trabajos en los que se aborda la integración de la HCE con SADC destacan por la falta de una API estándar para acceder a los datos de la HCE [120, 121]. A pesar de ello, los datos

que conforman las entradas de los SADC, y métodos para extraer estas entradas de la HCE, pueden definirse a través de modelos estándar de HCE (típicamente arquetipos) [6, 16, 91, 122]. Así, la combinación de esta idea con el enfoque de HCE aumentada puede conducir a una explotación más precisa y eficiente de los datos de HCE para usos secundarios, dado que los datos de HCE original y de fuentes externas pueden procesarse a partir de la HCE aumentada (o de una abstracción de la misma).

En este trabajo se han explotado las herramientas de web semántica existentes para descubrir relaciones equivalentes entre los valores codificados de la HCE y los términos de la nube LODC. Sin embargo, la mayoría de definiciones de esquemas semánticos en la LODC no están bien formadas en el dominio de LOD [123], al haber sido generadas automáticamente mediante sistemas de rastreo de enlaces, se dificulta enormemente la identificación automática de nuevos contenidos semánticamente relevantes. La calidad de estos conjuntos de datos LOD en términos de integridad, consistencia, concisión e interrelación sigue siendo un reto [124]. Los descriptores de vocabulario de conjuntos de datos entrelazados VoID, permiten identificar los conjuntos de datos LOD públicos cuando no son proporcionados por la organización que soporta la ontología. No obstante, no todos los conjuntos de datos publicados proporcionan descriptores VoID, y en los que sí se proporcionan, la información relevante (por ejemplo, el prefijo de la URL de los elementos de datos en el conjunto de datos LOD de destino) está distribuida de manera heterogénea en las diferentes propiedades de cada descriptor VoID (es decir, la información relevante no siempre se encuentra ubicada en la misma sección de los descriptores VoID para todos los conjuntos de datos LOD), lo que sugiere que se están utilizando de manera incorrecta o incompleta. Para promover la automatización del proceso de diseño de la HCE aumentada, se ha implementado un algoritmo que asiste en el proceso de identificar la ubicación adecuada del slot para el arquetipo de aumento en el arquetipo de HCE aumentada. El algoritmo se basa en la estructura sintáctica del arquetipo inicial, y considera las relaciones semánticas tratando de identificar una posición adecuada cerca de la EDI. Por ejemplo, un elemento hermano de la EDI, bajo el mismo componente padre en el arquetipo.

Aunque la solución de HCE aumentada propuesta requiere algunas tareas de modelado manuales para definir aumentos de HCE, una vez definidos, el proceso de utilizarlos para aumentar extractos de HCE es completamente automático, sin restricciones ni intervenciones manuales.

Capítulo 4. Informe radiológico estructurado enriquecido con sistemas de ayuda a la decisión clínica

4.1. Introducción

El IRE es una herramienta fundamental para la comunicación entre el radiólogo, el clínico y el paciente. Es el documento que integra el resultado final del proceso asistencial del radiólogo con el paciente, por lo que debe contener la información de forma estructurada, concisa y precisa. El informe debe ser comprensible tanto para el paciente como para el clínico, con el fin de tomar decisiones eficaces. Por ello, el IRE incluye datos del estudio redactados de forma estandarizada, en campos separados y tabulados para que puedan ser visualizados y explotados posteriormente. Los IRE deben estar bien organizados y contener, al menos, los siguientes apartados [125]: (I) datos demográficos del paciente. (II) Resumen de la información clínica relevante. Si no hay información, debe indicarse de manera explícita. (III) Descripción de la técnica de exploración, incluyendo metodología, secuencias, medio de contraste, y posibles limitaciones en la calidad de la imagen. (IV) Descripción detallada de los hallazgos, respondiendo a la justificación que motivó la exploración e incluyendo estudios comparativos, con fechas, siempre que sea posible. (V) Conclusión. Resumen de los hallazgos más relevantes y recomendación. Nótese que el Informe de resultados de pruebas de imagen, que aquí llamamos informe radiológico, es uno de los informes del CMDIC. En esta tesis trabajamos con el IRE, que contiene al anterior, es más sofisticado y aporta ventajas en cuanto a la normalización y calidad del informe radiológico.

El IRE introduce mejoras significativas en el informado radiológico, en contraposición con el informe radiológico dictado tradicional [55–57, 59–63]: (I) las plantillas de IRE específicas para cada enfermedad pueden mejorar la claridad y calidad de los informes, y garantizar en la práctica un uso coherente de las terminologías clínicas. (II) Los IRE basados en listas de comprobación o “checklists” pueden reducir los errores de diagnóstico, así como los errores gramaticales de los sistemas digitales de reconocimiento de voz. (III) Los IRE garantizan una documentación completa del informe radiológico, y tienen un impacto positivo en la investigación radiológica al facilitar la extracción de datos de manera estructurada. (IV) Los IRE pueden ayudar a promover la medicina basada en la evidencia al integrar los resultados de los SADC en los informes radiológicos. Y, por último, (V) en el contexto de la asistencia sanitaria centrada en el paciente, es necesario que los informes puedan ser leídos por profesionales sanitarios de diferentes especialidades e incluso por profesionales no sanitarios

(el paciente y sus cuidadores). El IRE puede utilizarse para generar contenidos en lenguaje natural de acuerdo con el nivel de comprensión del lector al que va dirigido. Para lograr las mejoras descritas, las soluciones actuales que implementan plantillas para la elaboración de informes radiológicos se centran en promover los IRE más allá del tradicional informe radiológico no estructurado basado en el dictado [64, 126], integrar los sistemas de puntuación cualitativa radiológica en los IRE [65–67], y garantizar una integración completa del IRE con la HCE utilizando estándares de sistemas de información sanitarios para HCE e imagen médica, así como el perfil de integración IHE MRRT de gestión de plantillas de IRE [68–71, 127].

A partir del estudio del estado del arte del IRE y su implantación en la práctica radiológica, y con la colaboración del comité de expertos en radiología del Hospital Universitari i Politècnic La Fe de Valencia, se han identificado aspectos de mejora que pueden mitigar algunas de las limitaciones actuales para la implantación del IRE en la práctica clínica radiológica (ver capítulo 2): (I) facilitar la transición de los informes convencionales a la cumplimentación de formularios electrónicos con un enfoque de lista de comprobación rápido y ágil. (II) Proporcionar funcionalidades para generar extractos de lenguaje natural a partir de los contenidos del informe, asegurando un texto de diagnóstico uniforme y adaptado a la audiencia (normalmente el médico destinatario o el paciente y sus cuidadores) y, al mismo tiempo, facilitando la integración con el PACS/RIS de forma similar a los sistemas convencionales de dictado y reconocimiento de voz. (III) Evitar que se pase por alto información relevante y minimizar los errores en la introducción de información, mediante mecanismos de validación de campos y reglas lógicas que automaticen el cálculo de valores relevantes (por ejemplo, el volumen de una lesión), garanticen el cumplimiento de los sistemas de puntuación cualitativa radiológica (por ejemplo, el sistema de puntuación PI-RADS⁵ [46]), y oculten o muestren dinámicamente campos del formulario de IRE para reducir la interrupción del patrón de búsqueda visual a lo largo del mismo. (IV) Poner a disposición del radiólogo los resultados de SADC integrados en el IRE en el momento de la evaluación del caso.

En este capítulo se plantea una solución de IRE enriquecido con información cuantitativa de biomarcadores de imagen, que implementa los aspectos de mejora mencionados. Específicamente, los requisitos del sistema de IRE enriquecido son: (I) integrar en el IRE información cualitativa del radiólogo (valoración y diagnóstico), con información cuantitativa obtenida a partir de SADC basados en biomarcadores de imagen. (II) Promover el uso del IRE a través de formularios interactivos, que fomentan mantener la atención en los campos relevantes del informe y reducir potenciales errores de informado mediante el uso de listas

⁵ PI-RADS: sistema radiológico de puntuación cualitativo para la evaluación del cáncer de próstata.

desplegables. (III) Permitir la definición de validadores de campos y reglas lógicas, para calcular parámetros y puntuaciones automáticamente y/o esconder o mostrar campos del formulario de IRE (según corresponda) conforme se rellena. (IV) Facilitar la integración del IRE con el PACS y/o el RIS, de forma similar a los sistemas de reconocimiento de dictado radiológico convencionales. (V) Permitir exportar el contenido del IRE en extractos de texto libre en lenguaje natural en diferentes idiomas. (VI) Ser compatible con las plantillas publicadas por la iniciativa Radreport [74] de la sociedad de radiología norteamericana RSNA (Radiological Society of North America). (VII) Cumplir con las especificaciones del perfil de integración IHE MRRT para la gestión de plantillas de IRE. Los dos últimos requisitos van dirigidos a soportar plantillas de IRE convencionales, como las de la iniciativa Radreport, además de plantillas de IRE enriquecido. En este contexto, se entiende por plantillas de IRE enriquecido aquellas que soportan la integración de resultados de SADC basados en biomarcadores de imagen y sistemas de puntuación cualitativos de guías clínicas radiológicas, además del resto de funcionalidades.

En este capítulo se propone una solución de IRE diseñada específicamente para cumplir con todos los requisitos mencionados anteriormente, y se describe un caso de uso del sistema con plantillas de IRE aprobadas por el TLAP de la RSNA, comité de expertos que valida y garantiza que las plantillas de Radreport incluyen solo información relevante para cada escenario clínico. En el caso de uso se presenta además una plantilla de IRE de próstata enriquecido con resultados de SADC, que cuenta con el sistema de puntuación cualitativo radiológico PI-RADS 2.1 y campos para embeber información de biomarcadores de IRM multi paramétrica de próstata. Concretamente, los campos para incluir resultados de biomarcadores de imagen contienen las posibles lesiones de próstata identificadas por el biomarcador y sus correspondientes valores de ADC, K^{trans} , k_{ep} , y v_e [50].

Las siguientes secciones en este capítulo presentan, por una parte, el diseño metodológico del sistema de IRE enriquecido con SADC a nivel de proceso y arquitectura de software (sección “4.2. Metodología”) y, por otra, una posible implementación de dicho sistema con tecnologías concretas, incluyendo la implementación del caso de uso de IRE de IRM próstata enriquecido con resultados de biomarcadores de imagen (sección “4.3. Resultados”). Finalmente, los resultados obtenidos, así como su comparación con otras soluciones existentes del estado del arte, son analizados en la sección “4.4 Discusión”.

4.2. Metodología

El sistema de IRE enriquecido ha sido diseñado de forma modular, permitiendo la creación de nuevas plantillas de IRE que pueden importarse en el sistema. El usuario (normalmente el radiólogo) deberá seleccionar la plantilla de IRE adecuada para la evaluación de cada caso.

El IRE escogido se mostrará en un formulario electrónico con campos y validadores específicos. El diseño modular proporciona flexibilidad para la personalización de plantillas de IRE según las necesidades de cada departamento de radiología. Esta característica facilita la creación de plantillas con un alto grado de especialización. Es posible diseñar una plantilla específica para cada patología, o incluso combinar el diagnóstico de diferentes patologías en una sola plantilla. Esto es especialmente útil para informar casos de pacientes con múltiples problemas. Por último, es importante destacar que el tiempo de informado varía considerablemente en función de diferentes factores, pudiendo llevar entre cinco y noventa minutos dependiendo de la patología a informar, el área anatómica, la complejidad de la plantilla de IRE y la experiencia del radiólogo que informa el caso.

Nótese que una plantilla de IRE representa un concepto análogo al de arquetipo de HCE. En ambos casos, son representaciones abstractas de conceptos clínicos que sirven para recoger y compartir información de salud de forma homogénea y normalizada entre sistemas que interpreten la misma norma (por ejemplo, la norma ISO13606 o la norma DICOM). De esta manera, es posible modelar un arquetipo de HCE que represente el/los mismo(s) concepto(s) clínico(s) que una plantilla de IRE, dando lugar a un arquetipo de IRE. Esta idea se desarrolla más adelante en el capítulo 5.

La implementación de una plantilla de IRE consiste en:

- (I) Un diseño de plantilla que defina la estructura del formulario de IRE con los campos que deben rellenarse.
- (II) Validadores de campos y condiciones, mediante reglas lógicas, que definen el comportamiento del formulario de IRE. Establece cuándo mostrar u ocultar campos específicos en función de la información proporcionada (por ejemplo, si se indica un hallazgo determinado, se muestran al usuario campos adicionales para incluir los detalles de ese tipo de hallazgo), cuándo calcular automáticamente un valor específico (por ejemplo, el volumen de la lesión puede calcularse a partir de la delimitación de la región de interés) o cuándo advertir a los usuarios en caso de que falte información necesaria o se proporcionen datos erróneos (por ejemplo, valores fuera de rango o tipos de datos incorrectos).
- (III) Especificaciones en lenguaje natural para convertir el contenido del IRE en extractos de lenguaje natural. Consisten en un conjunto de frases en diferentes idiomas relacionadas con cada campo de la plantilla del IRE. A medida que el radiólogo rellena el informe, el sistema incorpora la información de cada campo a estas frases y las combina para construir el extracto en lenguaje natural.
- (IV) Estructura de datos para contener la información dinámica de la plantilla (en adelante “objeto padre”), incluyendo información cualitativa, cuantitativa, y frases en lenguaje natural generadas a partir de los datos introducidos en la plantilla.

El diseño de la plantilla incluye un conjunto de variables llamadas "variables de enlace" que son utilizadas por las reglas lógicas para identificar los datos de cada fuente original (datos demográficos del paciente procedentes de la cabecera DICOM, datos cualitativos procedentes de la valoración radiológica, o datos cuantitativos resultantes del procesamiento de biomarcadores de imagen). Las reglas lógicas gestionan el comportamiento de la plantilla en función de los valores de las variables de enlace en cada momento. Cuando el usuario introduce datos en el informe, los datos son almacenados en las variables de enlace. Las reglas lógicas evalúan los datos y modifican dinámicamente la plantilla si es necesario. Las variables de enlace están contenidas en un objeto padre que contiene tres secciones principales: información cualitativa, información cuantitativa y frases en lenguaje natural. El radiólogo introduce manualmente, mediante escritura con el teclado, la información cualitativa durante el informado del caso. El proceso consiste en rellenar un formulario construido de acuerdo a la plantilla de IRE. El formulario contiene los campos especificados en la definición de la plantilla y aplica la evaluación de reglas lógicas y validadores de campos en tiempo real, durante el informado del caso. La información cualitativa introducida por el radiólogo y los valores calculados automáticamente por la plantilla, de acuerdo con los métodos de puntuación de las guías radiológicas para la valoración de lesiones, se almacenan en la sección de información cualitativa. Cuando se dispone de resultados de SADC cuantitativos (es decir, obtenidos de forma automática mediante el procesado de la imagen y sin la intervención del radiólogo), el sistema los incorpora automáticamente al formulario de IRE, en las variables de la sección de información cuantitativa. Estas variables son parte del formulario de IRE y el radiólogo puede eliminar o modificar la información cuantitativa si no está de acuerdo con los resultados proporcionados por el SADC. Durante el informado del caso, se van generando frases en lenguaje natural a medida que el radiólogo introduce o modifica datos en el formulario de IRE. Estas frases se almacenan en las variables de enlace, en la sección de sentencias en lenguaje natural del objeto padre. Al final del informado del caso, estas sentencias se combinan para construir el extracto de texto en lenguaje natural del IRE.

El sistema incluye una interfaz DICOM SCP para recibir imágenes médicas (secuencias nativas) y resultados de cuantificación (mapas paramétricos con valores cuantitativos de la imagen) como objetos DICOM enviados desde el PACS y/o las modalidades de adquisición. La recepción de imágenes y resultados de cuantificación en formato DICOM permite la extracción automática de datos del paciente de las cabeceras DICOM, y la inclusión de dichos datos en el formulario del IRE. El proceso completo se resume en la figura 22. Los datos cualitativos y cuantitativos se proporcionan al formulario de la plantilla (paso 1), cuyas variables de enlace se vinculan al objeto padre (paso 2). Cada vez que estas variables son modificadas, se desencadena la evaluación de validadores de campos y reglas lógicas (paso

3). La evaluación puede conducir a cambios en los valores actuales de las variables de enlace (por ejemplo, valores calculados automáticamente, paso 4a) o a cambios en el formulario de IRE (por ejemplo, ocultar o mostrar campos específicos, paso 4b). A continuación, se cargan las especificaciones en lenguaje natural (paso 5) y se generan frases en lenguaje natural según los valores actuales de las variables de enlace (paso 6). Por último, las frases en lenguaje natural se combinan para generar el extracto del informe en lenguaje natural (paso 7).

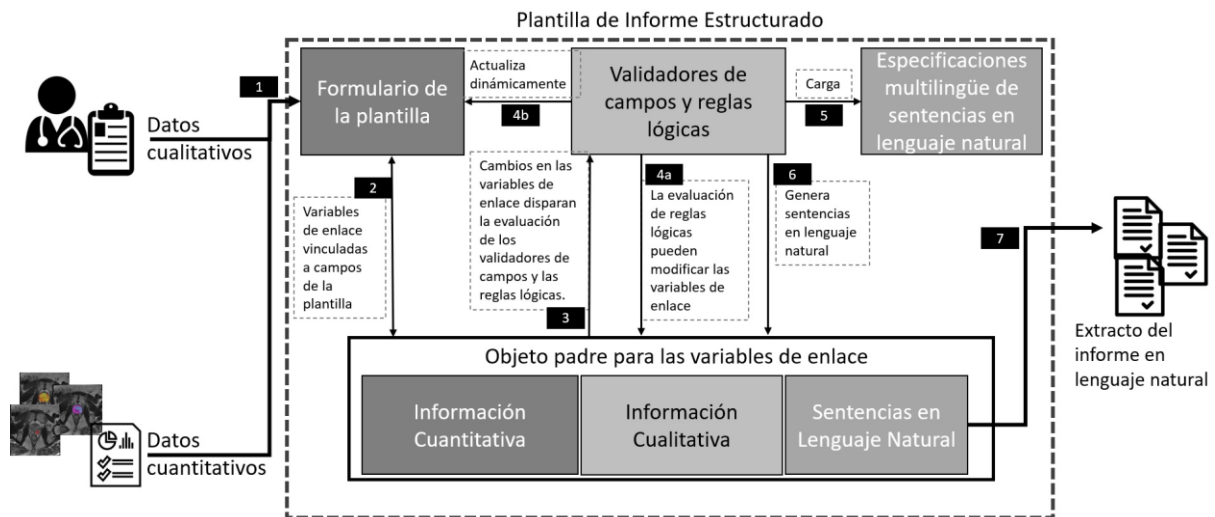


Figura 22. Flujo de trabajo del informado de un caso con el sistema de IRE enriquecido. El proceso se describe en términos de los componentes de la plantilla de IRE (formulario de la plantilla, validadores de campos, reglas lógicas y especificaciones de lenguaje natural), y del objeto padre que contiene los valores de las variables de enlace (información cuantitativa, cualitativa, y sentencias en lenguaje natural).

El diseño contempla la integración del sistema de IRE enriquecido con los sistemas de información de salud utilizando estándares de informática médica. De acuerdo con las especificaciones del perfil de integración IHE MRRT, el enfoque modular del sistema permite la definición de plantillas de IRE empleando el lenguaje de maquetación de hipertexto HTML5 para el diseño de la plantilla, y establece mecanismos de transporte para consultar, recuperar y almacenar las plantillas de IRE. Al finalizar el informado, el contenido del IRE se exporta en formato DICOM-SR para poder ser explotado mediante los sistemas de información PACS y RIS. Por otro lado, se genera el extracto en lenguaje natural y se ofrece al usuario la opción de enviarlo en un mensaje HL7 v2, de forma similar a los sistemas convencionales de dictado radiológico. Nótese que el sistema solo soporta protocolo HL7 v2 para el envío de informes, pero implementa comunicaciones DICOM tanto para recibir como para enviar objetos DICOM. La figura 23 muestra un ejemplo de la solución de IRE desarrollada en esta tesis, en la que la información procedente de diferentes fuentes (diagnóstico del radiólogo, biomarcadores de imagen, sistemas de puntuación cualitativa radiológica y valores calculados automáticamente) se combina en un extracto de texto en lenguaje natural.

Diagnóstico radiológico

Extensión extracapsular	<input type="text" value="Sí"/>
Afectación del paquete vasculonervioso periprostático	<input type="text" value="Sí"/>
Afectación a las vesículas seminales	<input type="text" value="Sí"/>
Cuello de la vejiga	<input type="text" value="Normal"/>
Esfínter distal	<input type="text" value="Normal"/>
Otras estructuras pélvicas	<input type="text" value="No"/>
Ganglios linfáticos	<input type="text" value="N0 (no sospecha)"/>
Hueso	<input type="text" value="Sin lesiones de tipo agresivo"/>

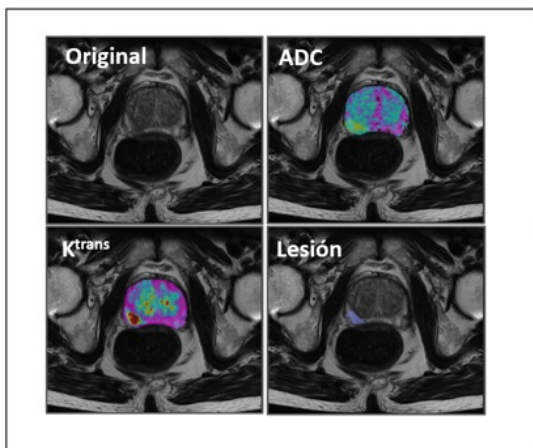
Valores calculados automáticamente

Información precalculada a partir de la segmentación automática de la próstata

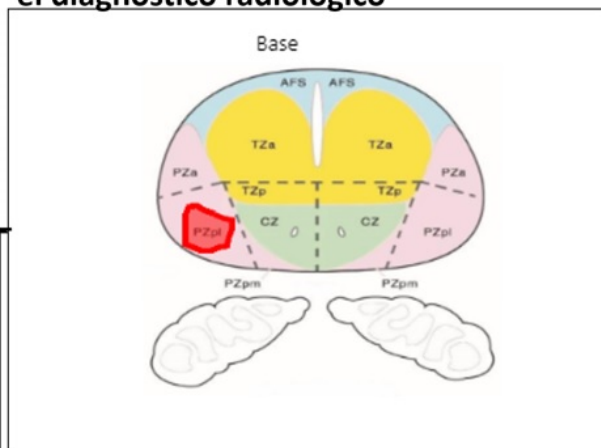
Tamaño: x x mm

Volumen: mL

Biomarcadores de imagen



Sistemas de puntuación cualitativa para el diagnóstico radiológico



Extracto del informe en lenguaje natural

<p>La próstata presenta una apariencia heterogénea. El tamaño de la próstata es 56 x 45.6 x 45 mm. El volumen de la próstata es 68.54 mL. Hay una lesión en la próstata. La puntuación PI-RADS fue calculada con las secuencias DWI, DCE y TE.</p> <p>Factores pronósticos: Presencia de extensión extracapsular. El haz neurovascular periprostático está afectado. Las vesículas seminales están afectadas. No se aprecia invasión en el cuello de la vejiga. No se aprecia infiltración en el esfínter distal. Otras estructuras pélvicas están afectadas.</p>	<p>Ganglios linfáticos: N0. No hay cáncer en ganglios linfáticos regionales.</p> <p>Hueso: Los huesos no muestran lesiones agresivas.</p> <p>Clasificación TNM: Tumor primario (T): T4 tumor fijo o tumor que invade estructuras adyacentes que no son las vesículas seminales: cuello de la vejiga, esfínter externo, recto, músculos elevadores y/o pared pélvica. Ganglios linfáticos (N): N0. No hay cáncer en los ganglios linfáticos regionales Metástasis (M): M1a. Ganglios linfáticos no regionales</p> <p>Conclusiones: Encontrada una lesión significativa.</p>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Figura 23. Ejemplo de cómo la información relevante del informe procedente de distintas fuentes (diagnóstico del radiólogo, biomarcadores de imagen, sistemas de puntuación cualitativa radiológica y valores calculados automáticamente) se combina en un extracto de texto en lenguaje natural.

4.3. Resultados

El proceso descrito en la figura 22 se ha implementado en un sistema basado en web, para facilitar la integración del IRE enriquecido en el flujo de trabajo radiológico al tiempo que

permitir la interacción con resultados de SADC. El sistema se ha desarrollado utilizando la pila MEAN [128], compuesta por la base de datos MongoDB para la capa de persistencia, el servidor de aplicaciones web ExpressJS para albergar los servicios de backend, el framework Angular para la interfaz web y NodeJS para la implementación de los servicios de backend. Además, se han empleado herramientas basadas en Java para las operaciones relacionadas con los estándares DICOM y HL7. En concreto, se ha utilizado la librería dcm4che toolkit para la gestión de objetos y comunicaciones DICOM (es decir, para convertir el informe a objeto DICOM-SR y enviarlo a PACS), y la librería HAPI para enviar el informe a través de mensajes HL7 v2.

Las plantillas de IRE se importan en el sistema como módulos independientes o plug-ins. Cada plantilla de IRE se implementa con HTML5 para el diseño de la plantilla, JavaScript para los validadores de campos y las reglas lógicas, y objetos JSON para definir las especificaciones del lenguaje natural en varios idiomas. La figura 24 ilustra una porción de estos archivos de una plantilla de IRE de IRM de próstata, así como el correspondiente formulario y el extracto de lenguaje natural generado. La metodología empleada para generar validadores de campo, reglas lógicas y extractos de lenguaje natural se aplica también para la implementación de sistemas de puntuación cualitativos en plantillas de IRE.

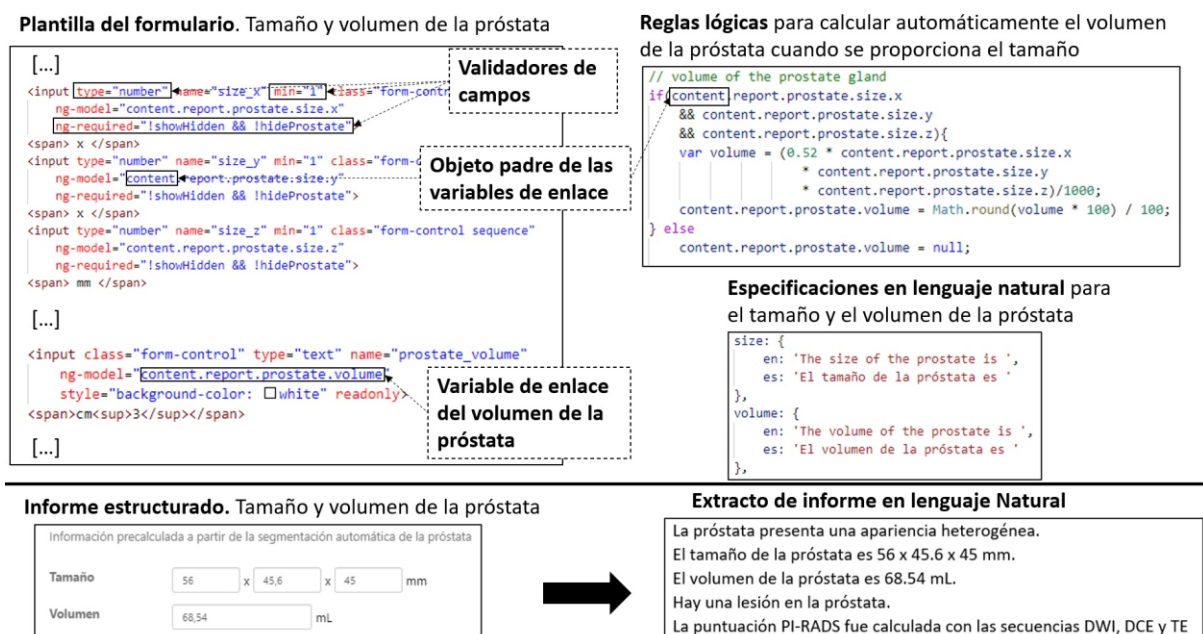


Figura 24. Una sección de la plantilla de IRE de IRM de próstata relacionada con el tamaño y el volumen de la próstata. En la parte superior de la figura se muestra el diseño de la plantilla, los validadores de campos, una regla lógica y las especificaciones de texto en lenguaje natural. En la parte inferior, se muestra el formulario de IRE y el extracto de texto en lenguaje natural generado. El volumen de la próstata se calcula automáticamente mediante la regla lógica en función del tamaño de la próstata.

Los validadores de campo y las reglas lógicas son útiles para advertir a los usuarios cuando se proporcionan valores incorrectos, o cuando falta la información requerida. Permiten definir

campos como obligatorios u opcionales, y limitarlos a un tipo de datos específico y/o a un rango de valores concreto. Además, las reglas lógicas pueden emplearse para ocultar o mostrar los campos del informe en función de la información proporcionada por el usuario. La figura 25 muestra ejemplos de cómo estos elementos restringen los valores posibles, advierten a los usuarios cuando se introducen valores incorrectos y ocultan o muestran los campos en función de los valores proporcionados.

1

2

3

Identificador	Tamaño (mm)	Lado	Región
1	<input type="text"/>	<input type="text"/>	<input type="text"/>

Por favor, completa este campo

Por favor, selecciona una opción

Por favor, selecciona una opción

Figura 25. Validadores de campos y reglas lógicas que cambian dinámicamente el formulario del IRE y muestran u ocultan los campos según las entradas del usuario. Transición 1: El campo "Número de ganglios linfáticos" aparece cuando se selecciona "N+ (Sospecha de metástasis)" en el campo "Ganglios linfáticos". Transición 2: Aparece una tabla, para registrar información detallada de los ganglios linfáticos, cuando se selecciona "1" en el campo "Número de ganglios linfáticos". Transición 3: Se muestran advertencias al usuario cuando se intenta finalizar el informado con valores ausentes y/o incorrectos.

El tiempo necesario para implementar una plantilla de IRE varía considerablemente en función de la complejidad de la misma. Es posible implementar una plantilla sencilla en menos de una hora, o dedicar varios días a la definición e implementación de una plantilla compleja. Las plantillas pueden registrarse en el sistema a través de un panel de administración, que también sirve para habilitarlas o deshabilitarlas. Para registrar una nueva plantilla de IRE, es necesario proporcionar la definición del diseño de la plantilla, los validadores de campos y las reglas lógicas, y las especificaciones de texto en lenguaje natural. Estos componentes pueden proporcionarse en un único archivo HTML para cumplir con el perfil de integración IHE MRRT, o como archivos separados comprimidos en zip. El sistema requiere información adicional para registrar una nueva plantilla: título, descripción y aplicación clínica, para poder indexar las plantillas y facilitar su búsqueda y acceso. La figura 26 muestra los componentes web desarrollados para soportar el ciclo de vida de las plantillas de IRE, que cumple con las

especificaciones del perfil de IHE MRRT, proporcionando funcionalidades para almacenar, consultar y recuperar plantillas de IRE HTML5. La compatibilidad con el perfil de integración IHE MRRT se logra utilizando el protocolo HTTP, tal y como especifica el marco técnico de IHE MRRT, para recuperar (“4.103 Retrieve Imaging Report Template [RAD-103]”), almacenar (“4.104 Store Imaging Report Template [RAD-104]”) y consultar (“4.105 Query Imaging Report Template [RAD-105]”) plantillas. Para facilitar la integración con los sistemas PACS/RIS, el formulario de la plantilla permite exportar el informe como DICOM-SR o como texto plano (a través de un mensaje HL7 v2) al PACS/RIS.

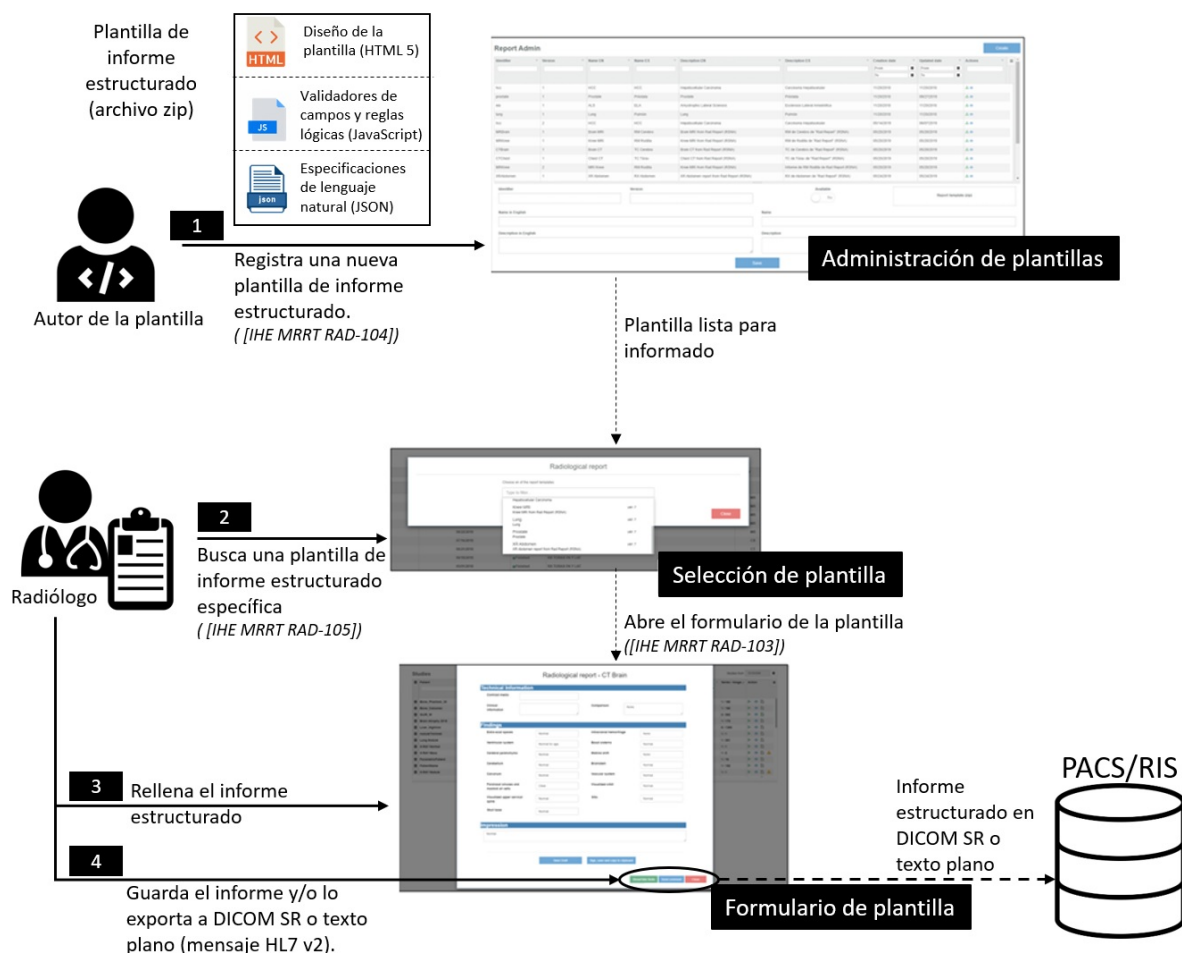


Figura 26. Ciclo de vida de la plantilla de IRE conforme al perfil de integración IHE MRRT. En primer lugar, el autor registra la plantilla de IRE en la interfaz de administración de plantillas. En este punto se proporciona la información contextual de la plantilla (título, descripción y aplicación clínica). A continuación, la plantilla está disponible para la elaboración de informes. Un radiólogo puede buscar y seleccionar la plantilla para rellenar el formulario con la información del diagnóstico. Por último, el radiólogo puede almacenar el informe y exportarlo como DICOM SR y/o como texto plano, de forma similar a los sistemas convencionales de dictado radiológico.

El sistema de IRE enriquecido almacena la información cuantitativa de biomarcadores de imagen en las variables de enlace, posibilitando la integración con plataformas de biomarcadores de imagen [12] y la incorporación de datos cuantitativos en el contenido del informe. Se ha llevado a cabo una implementación de este enfoque integrando el sistema de

IRE enriquecido con la plataforma de biomarcadores de imagen QUIBIM Precision® (QUIBIM, Valencia, España). El resultado es una herramienta avanzada de informado radiológico, que enriquece el IRE con resultados de SADC. Las variables de enlace para información cuantitativa están vinculadas a la salida de la plataforma de biomarcadores de imagen, facilitando embeber dinámicamente los resultados cuantitativos en el formulario del IRE para proporcionar apoyo a la decisión del radiólogo durante el informado.

Un ejemplo de dos de las secciones más relevantes de la plantilla de IRE de IRM de próstata, enriquecido con resultados de SADC basados en biomarcadores de imagen, se muestra en las figuras 27 y 28 (sistema de puntuación PI-RADS y resultados de SADC, respectivamente). En el ejemplo se observa un caso de próstata con una lesión dominante. En la sección para la puntuación PI-RADS 2.1 (figura 27), el radiólogo debe rellenar manualmente las puntuaciones en función del aspecto de la lesión en las secuencias T2, DWI y DCE, el diámetro y la zona de la lesión. La plantilla contiene listas desplegables con los valores de las posibles puntuaciones, para restringir los valores que el usuario puede introducir y evitar errores. La puntuación PI-RADS 2.1 final se calcula automáticamente según los valores introducidos. La guía radiológica PI-RADS 2.1 para la evaluación de lesiones de próstata establece un número máximo de cuatro hallazgos representativos, correspondientes con aquellos que tengan la puntuación PI-RADS 2.1 más alta, para diagnosticar las lesiones de próstata. Siguiendo esta recomendación, la plantilla de IRE de IRM de próstata permite gestionar hasta cuatro lesiones, facilitando al usuario herramientas para añadir, editar o eliminar lesiones dominantes mediante listas desplegables y un formulario en forma de tabla. Específicamente, la lista desplegable “Lesiones y valoración PI-RADS” (ver figura 27) sirve para establecer el número de lesiones dominantes y la lista “Lesión dominante” se utiliza para escoger la lesión dominante a mostrar en el esquema PI-RADS del formulario. Por su parte, el formulario en forma de tabla permite introducir las puntuaciones PI-RADS 2.1 de cada lesión dominante, proporcionando para ello listas desplegables y campos con tipos de datos y rangos de valores restringidos.

En el caso del ejemplo, la plataforma de biomarcadores ha detectado una posible lesión dominante (DL 1) en la adquisición y ha cuantificado el volumen de lesión, y los valores de v_e , ADC, K^{trans} , y k_{ep} en la región donde se ha identificado la lesión. El sistema ha embebido los valores cuantificados de la lesión DL 1 en la sección de información cuantitativa de las variables de enlace, poniendo a disposición del usuario la información cuantitativa en el formulario de IRE (figura 28). El radiólogo puede modificar o eliminar una lesión mal identificada manipulándola o borrándola en el visor (parte izquierda de la figura 28). La plantilla de IRE para IRM de próstata también permite añadir al IRE imágenes clave de los algoritmos de cuantificación, como se muestra en la parte inferior derecha de figura 28, que

deben ser seleccionadas manualmente por el radiólogo en función de su interés potencial para el diagnóstico y el facultativo destinatario del IRE.

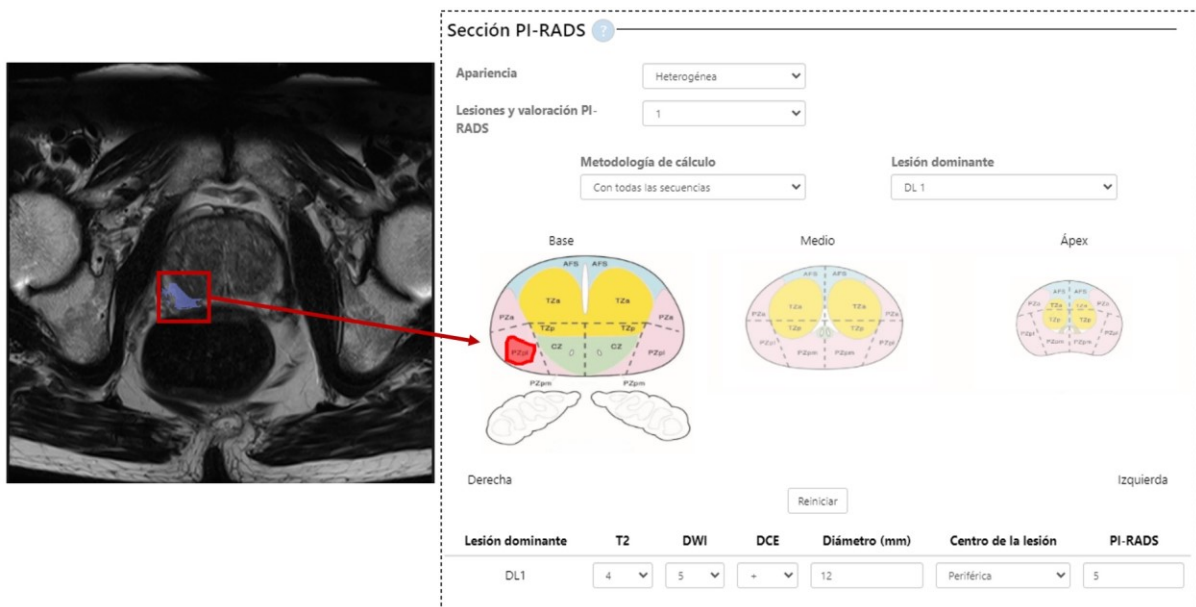


Figura 27. IRE de IRM de próstata, sección PI-RADS. La plantilla incluye herramientas para la evaluación de la lesión de próstata según la guía radiológica PI-RADS 2.1. El radiólogo evalúa el estadio de T2, DWI y DCE, según los criterios de puntuación establecidos en la guía. Además, es posible delimitar la lesión sobre el diagrama de sectores de la próstata definido en dicha guía.

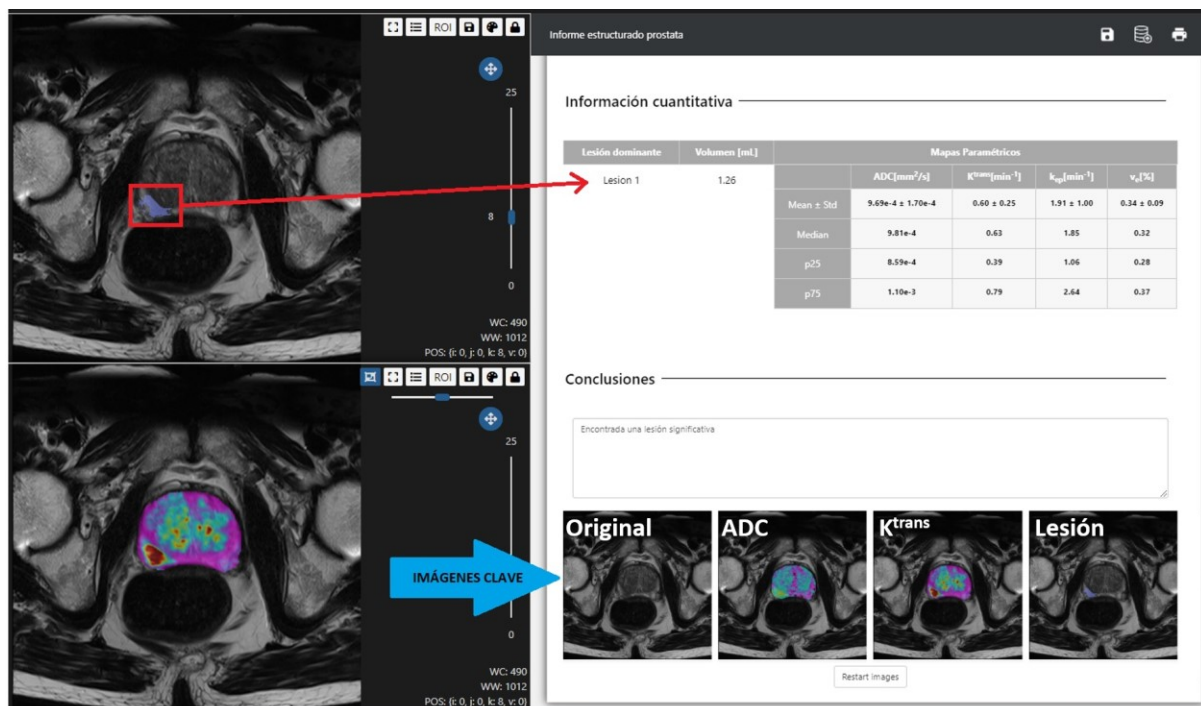


Figura 28. IRE de IRM de próstata, integrado con una plataforma de biomarcadores de imagen. La plantilla incluye variables de enlace para los datos cuantitativos de los biomarcadores de imagen (sección “Información cuantitativa”). El sistema detectó una posible lesión (Lesión 1) y sus valores de K^{trans} , V_e , K_{ep} y ADC se han embebido automáticamente en el formulario del IRE. Además, la plantilla incluye variables de enlace para las imágenes clave. Durante el informado, el radiólogo puede seleccionar un conjunto de imágenes clave relevantes para el diagnóstico, e incluirlas en el IRE.

Hasta la fecha de escritura de esta tesis, se han implementado e integrado en el sistema 24 plantillas de IRE, algunas de las cuales se detallan en el Anexo I. Entre ellas, 17 son plantillas aprobadas por el TLAP y 7 son plantillas de IRE enriquecido personalizadas. Estas últimas incluyen resultados de biomarcadores de imagen y de sistemas radiológicos de puntuación cualitativos PI-RADS, BI-RADS y LI-RADS. Las plantillas de IRE enriquecido han sido desarrolladas y adaptadas a los requisitos específicos de los departamentos radiológicos que utilizan el sistema. Por el momento, el sistema está instalado en 37 hospitales e instituciones sanitarias (ver lista completa en el Anexo II) para guiar el proceso de informado y homogeneizar el informe radiológico. Un total de 949 radiólogos de diferentes centros se han registrado en este sistema y han informado 21.853 casos.

4.4. Discusión

Investigaciones disponibles en la literatura sobre sistemas de plantillas de IRE basadas en el consenso [64, 69, 70], plantean soluciones que carecen de reglas lógicas avanzadas para evitar errores y mantener la coherencia de los informes. Los sistemas de dictado convencionales a menudo incorporan reglas lógicas básicas de tipo “si/si no” (“if/else”), sin embargo, las reglas lógicas de la solución presentada en este trabajo se implementan directamente con el lenguaje de programación JavaScript, facilitando embeber en las plantillas sofisticados algoritmos como el cálculo de la puntuación PI-RADS 2.1 o la generación de extractos en lenguaje natural para asistir al radiólogo en el proceso de informado. Además, a diferencia de nuestra metodología, las soluciones avanzadas que proporcionan calculadoras para sistemas de puntuación radiológica [65], y marcos para la elaboración de informes asistidos por ordenador y el apoyo a la toma de decisiones [66, 67], carecen de capacidad para integrar resultados cuantitativos de biomarcadores de imagen con los datos cualitativos del IRE en tiempo de informado.

La generación de extractos de texto en lenguaje natural y en diferentes idiomas, a partir del contenido del informe, sirve para adaptar el informe al público objetivo que lo ha de recibir (normalmente, el médico destinatario o el paciente y sus cuidadores). El procedimiento para crear una plantilla de IRE dirigida a una audiencia específica, o en múltiples idiomas, o para una audiencia específica en múltiples idiomas, es similar en todos los casos y dependerá del uso previsto de la plantilla de IRE. Este trabajo presenta un ejemplo de una plantilla que soporta múltiples idiomas, pues en la literatura se señala la falta de soporte para ello como una de las limitaciones actuales del IRE. Sin embargo, se puede aplicar este mismo mecanismo a una plantilla cuyo objetivo sea proporcionar un extracto del informe en lenguaje natural adaptado para un especialista, un abogado, o un paciente y sus cuidadores sin conocimientos médicos (por ejemplo, un familiar). En el caso de uso presentado en esta tesis,

la plantilla contiene una variable que almacena el idioma preferido por el usuario, y utiliza esta variable para indexar las sentencias en lenguaje natural, ubicadas en la definición de la plantilla, y combinarlas para generar una versión del informe expresada en lenguaje natural. Es posible aplicar este mismo método para otros objetivos. Por ejemplo, una plantilla que incluye una lista desplegable para elegir la audiencia objetivo (ej. paciente o especialista), puede utilizar el valor de la variable de enlace de dicha lista para indexar sentencias en lenguaje natural definidas en la plantilla. Así, el extracto en lenguaje natural cambiará de acuerdo con la audiencia objetivo seleccionada en la lista desplegable. Para contrarrestar el efecto de la curva de aprendizaje del estilo de informado basado en el IRE, que puede repercutir negativamente en el flujo de trabajo y la productividad de la radiología [55], se incluyen sistemas de puntuación cualitativa radiológicos y de biomarcadores de imagen, que pueden ahorrar tiempo en el proceso de toma de decisiones, y así animar a los radiólogos a empezar a utilizar el IRE para informar casos. Por otro lado, la interrupción del patrón de búsqueda visual puede mitigarse mostrando y ocultando los campos del formulario del informe de forma dinámica en función de los valores proporcionados, centrando la atención del usuario en las secciones relevantes del formulario.

El sistema de IRE enriquecido gestiona plantillas programadas en HTML5, que contienen objetos JSON y código JavaScript en un único archivo html, cumpliendo con el perfil de integración de IHE MRRT y, al mismo tiempo, aprovechando las ventajas de los lenguajes de script para formularios dinámicos e interactivos. Además, el sistema puede importar plantillas de IRE de la iniciativa Radreport de la RSNA. Para modelar nuevas plantillas compatibles, la RSNA proporciona el editor de plantillas de IRE T-Rex [129]. Se requieren conocimientos de programación HTML y JavaScript para enriquecer las plantillas con reglas lógicas, validadores de campos, sistemas de puntuación y/o resultados de biomarcadores de imagen. Típicamente, las plantillas de IRE las crean equipos multidisciplinares con experiencia en la práctica radiológica y en lenguajes de programación. Las reglas lógicas y los validadores de campos en las plantillas deben definirse para guiar y asistir el proceso de informado radiológico. Deben restringir los posibles valores de entrada del formulario de IRE para evitar errores, pero sin limitarlos excesivamente para no dificultar el informado de casos raros o atípicos. Una vez definidas, las plantillas de IRE pueden emplearse para la elaboración de informes sin necesidad de conocer estos lenguajes de programación. Solo requieren un navegador web para mostrarla al usuario y ser rellenas.

Para garantizar que se incluya la información pertinente y necesaria en los informes, y al mismo tiempo descartar información irrelevante, se han desarrollado plantillas de IRE a partir de plantillas de Radreport aprobadas por el TLAP y/o siguiendo las especificaciones de comités de expertos en radiología. La interfaz de administración (figura 26) sirve para personalizar el porfolio de plantillas de IRE de cada escenario clínico, e incluye un mecanismo

de versionado que permite trazar la evolución de las plantillas. Una vez que se notifican los casos, los informes se almacenan en una base de datos MongoDB, que ofrece flexibilidad para almacenar cualquier tipo de dato formateado como objeto JSON (objeto padre de las variables de enlace). Al almacenar los datos del diagnóstico en un formato estructurado, se fomenta el uso secundario de los informes radiológicos en investigación y radiómica [130], facilitando el uso de Big Data y de herramientas de Inteligencia Artificial (IA) para explotar los datos de los informes. Estos datos se pueden explotar con consultas a la base de datos MongoDB, por medio de herramientas de minería de datos e inteligencia de negocio (Business Intelligence, BI), o exportando la información de los informes en una hoja de datos estructurada para su posterior análisis. Otras funciones del sistema de IRE enriquecido son el uso de metadatos de plantillas de informes estructurados para buscar y clasificar plantillas, la capacidad de exportar informes estructurados como DICOM SR o texto plano y la integración con sistemas PACS/RIS utilizando protocolos de comunicación DICOM y mensajería HL7 v2. El mecanismo de versionado y el formulario de búsqueda de plantillas de IRE ayudan al usuario a identificar la plantilla más adecuada, para cada tipo de patología y área anatómica, de entre grandes repertorios de plantillas disponibles.

Capítulo 5. Discusión

Conceptualmente, la HCE aumentada consiste en el enriquecimiento de una HCE existente con contenidos adicionales (es decir, información y/o conocimiento) generados a partir de diferentes fuentes de información externas. En esta tesis, proponemos un enfoque preliminar para construir la HCE aumentada, homogeneizando el formato de las fuentes de información externas mediante su representación en forma de datos LOD, y utilizando dicha representación para aumentar las HCE con datos relacionados con la salud del paciente (a nivel de instancia) y/o datos relacionados con conceptos médicos (a nivel de esquema o modelo de información). Cabe destacar, que aunque el enfoque propuesto se basa en el uso de datos enlazados, puede aplicarse para enriquecer la HCE con contenidos adicionales derivados de fuentes externas, distintas a las de la web semántica. Para lo cual sería necesario diseñar transformaciones intermedias para construir una representación semántica de dichas fuentes. Asimismo, definimos las bases prácticas para expresar los datos de la web semántica como extractos normalizados de HCE. Esta solución proporciona un medio alternativo para salvar la brecha entre los estándares de sistemas de información sanitarios y la web semántica, transformando en extractos de HCE la información de la web semántica mediante los tres pilares básicos de la interoperabilidad semántica, definidos por la red europea de excelencia SemanticHealthNet: modelo de información, modelo de arquetipos y modelos de dominio (terminologías y/o ontologías) [23].

Uno de los elementos importantes de la HCE es el IRE, con potencial para integrar protocolos, observaciones, datos, sistemas de ayuda y guías. Este informe híbrido, cualitativo y cuantitativo, que en esta tesis hemos denominado IRE enriquecido con resultados de SADC, permite trabajar con una estructura predefinida en áreas complejas y relevantes. Es un enlace directo a los datos radiómicos, de los biomarcadores de imagen y sus imágenes paramétricas claves, soportando sistemas de clasificación automática y brindando apoyo a la toma de decisiones [131]. El resultado de la investigación nos ha llevado al desarrollo de un sistema de IRE enriquecido con resultados de SADC, que gestiona plantillas de IRE de acuerdo al perfil de integración de IHE MRRT. En este sistema, las plantillas de IRE incorporan mecanismos basados en validadores de campos y reglas lógicas para guiar el informado, calcular automáticamente determinados valores, embeber sistemas de puntuación de guías clínicas de radiología y/o resultados cuantitativos de SADC, disminuir las incertidumbres, evitar el uso de un lenguaje evasivo para minimizar responsabilidades, generar extractos del informe en lenguaje natural y reducir errores durante el proceso de informado. Nótese que esto último depende en gran medida de la calidad de la plantilla de IRE. Una plantilla bien organizada ayuda a reducir la tasa de errores, mientras que una plantilla imprecisa o mal

definida (por ejemplo, con reglas lógicas incorrectas o campos con información irrelevante) puede dificultar el informado e incrementar la tasa de errores.

Atendiendo a las definiciones de los términos aumentar⁶ y enriquecer⁷, consideramos más adecuado el concepto de HCE aumentada, dado que la HCE adquiere una mayor extensión de información al incorporar nuevos contenidos de fuentes externas, no necesariamente implícitos en la información inicial (por ejemplo, a partir del nombre de un fármaco de la medicación activa del paciente no es posible deducir, inferir o calcular sus contraindicaciones). Por otra parte, el verbo enriquecer define con más precisión el trabajo realizado sobre el IRE, puesto que se aumenta la cantidad de información que proporciona la imagen, derivada del procesado de la misma mediante SADC (por ejemplo, a partir de la imagen es posible deducir una potencial lesión y calcular su volumen). No obstante, el significado semántico de HCE enriquecida e IRE aumentado también representan una definición válida del resultado de nuestra investigación. En esta tesis, se emplean indistintamente ambos términos (enriquecer y aumentar) para referirse a la HCE aumentada y al IRE enriquecido.

En nuestro caso de uso de HCE aumentada, identificamos aumentos de HCE de potencial interés en el contexto de la HCE estándar utilizando arquetipos y recursos de la web semántica disponibles públicamente. A partir de estos aumentos, se ha diseñado un método y una implementación que permite aplicar dichos aumentos. Aunque este método fue diseñado en estrecha relación con el esquema de la HCE, los aumentos de HCE pueden ser modelados independientemente del extracto de HCE origen. Por lo tanto, es posible proporcionar un conjunto variado de aumentos de HCE a disposición de los especialistas, para que escojan el más conveniente para enriquecer extractos específicos de HCE. Los aumentos han sido diseñados para ser reutilizables. Esto significa que, una vez definido un aumento, es posible aplicarlo sobre más de una instancia de HCE de un determinado arquetipo⁸. Además, los aumentos de HCE también tienen potencial para enriquecer resultados de SADC. Para ello, una posibilidad es modelar una representación del IRE enriquecido en formato de HCE normalizada (arquetipo de IRE, en adelante). Posteriormente, aplicar el método de HCE aumentada sobre el arquetipo de IRE para obtener el contenido adicional relacionado con los resultados del SADC. El método de HCE aumentada fue diseñado con el objetivo de ser lo más automático posible. Sin embargo, teniendo en cuenta las limitaciones actuales de los esquemas semánticos (muchos de ellos han sido construidos mediante mecanismos automáticos de rastreo web, dando lugar a relaciones semánticas basadas en la estructura HTML en vez de en las relaciones conceptuales entre los datos) y

⁶ Aumentar: Dar mayor extensión, número o materia a algo [132].

⁷ Enriquecer: Aumentar en una mezcla la proporción de uno de sus componentes [133].

⁸ "Arquetipo inicial". Introducido en el capítulo 3.

descripciones de conjuntos de datos LOD (no todos los conjuntos de datos LOD cuentan con este tipo de descripciones, y no todas las descripciones de conjuntos de datos LOD tienen la información relevante en las mismas secciones), es necesario incluir algunas tareas de modelado manual en la definición de un aumento de HCE. El objetivo principal de la intervención manual es identificar los datos LOD relevantes y garantizar su coherencia semántica al embeberlos en el arquetipo aumentado. Esta labor debería ser realizada por expertos clínicos.

Por otra parte, el sistema de IRE enriquecido está preparado para las rutinas clínicas y el flujo de trabajo de los informes. Se ha implantado en 37 hospitales e instituciones sanitarias (lista completa disponible en Anexo II) y actualmente está en funcionamiento para la elaboración de informes radiológicos. Los hospitales e instituciones sanitarias desarrollan sus propias plantillas combinadas con biomarcadores de imagen utilizando este sistema. En el momento de escribir esta tesis, el sistema ha gestionado 21853 estudios de imagen informados por un total de 949 usuarios registrados de diferentes centros. Por defecto, el sistema incluye 17 plantillas de Radreport aprobadas por el TLAP, y 7 plantillas de IRE enriquecido con resultados de biomarcadores de imagen y calculadoras de sistemas de puntuación cualitativos BI-RADS, LI-RADS y PI-RADS (el Anexo I muestra un listado representativo de dichas plantillas). Su diseño modular permite que el sistema se instale de forma aislada o que se integre con una plataforma de biomarcadores de imagen. Nótese la diferencia entre plantillas de IRE (24 en total) y casos informados con formularios de IRE (21853). Una plantilla es la definición del conjunto de datos relevantes para el informado de una patología o área anatómica concreta, mientras que un caso informado es el resultado de rellenar un formulario de plantilla de IRE y emitir un informe con el diagnóstico de un caso específico.

En el contexto de esta tesis, se ha diseñado, implementado y puesto en producción un caso de uso de plantilla de IRE enriquecido con resultados de SADC, para el informado de IRM de próstata. La plantilla incluye campos para el diagnóstico, incluyendo el sistema de puntuación radiológico PI-RADS (información cualitativa) e información resultante de la cuantificación de biomarcadores de imagen de próstata (información cuantitativa). El éxito de esta contribución es la implementación, instalación y uso de una solución que cumple con los requisitos propuestos (ver capítulo 4), mostrando un impacto en el informado radiológico al permitir enviar informes en lenguaje natural al tiempo que almacena los datos del mismo en formato estructurado.

Existen potenciales sinergias entre la HCE aumentada con datos LOD y el IRE enriquecido con resultados de SADC. Integrar ambas soluciones puede añadir valor a la HCE aumentada, al permitir incorporar resultados de SADC y guías. Por otro lado, también abre la puerta a aumentar el IRE y/o los resultados de SADC con información potencialmente relevante de fuentes externas. Un sistema de información que combine ambas soluciones, sería

compatible con estándares de HCE e imagen médica, permitiría incorporar información radiológica cuantitativa y cualitativa a la HCE normalizada, y podría combinar la información de salud del paciente y los resultados de SADC para obtener información potencialmente relevante de fuentes externas.

La integración de los estándares de HCE e imagen médica es beneficiosa para el intercambio de información de salud entre sistemas de información sanitarios. El formato DICOM es el estándar para el almacenamiento y la transmisión de imágenes médicas. Cada archivo DICOM contiene abundante información de imagen. Sin embargo, debido a que la lectura de un DICOM implica la búsqueda de diccionarios de cabeceras DICOM y a que gran parte de las cabeceras son privadas, la legibilidad de un archivo DICOM es limitada. ISO13606, openEHR o HL7 son especificaciones estándar en informática médica para describir la gestión, almacenamiento, recuperación e intercambio de datos de salud en los registros de HCE. Uno de los retos más importantes que se plantean al desarrollo de estos modelos es el problema de la compartición de la información, especialmente en lo que respecta a la compartición de la información de imagen. Publicaciones en la literatura científica describen metodologías para integrar estas normas [31, 134, 135]. El planteamiento habitual consiste en transformar objetos del estándar DICOM a instancias de las normas de HCE. DICOM incluye un modelo de referencia para modelar plantillas de IRE, con un único tipo de componente, en DICOM-SR (el modelo de referencia DICOM-SR se describe en detalle en el capítulo 2). En contraste, los estándares de HCE basados en el modelo dual, proporcionan mayor diversidad de componentes en sus MR para modelar la información clínica. El modelo de referencia establece una jerarquía entre ellos que restringe el modelo de arquetipos. Dicha jerarquía, está diseñada para soportar la HCE completa. Mientras que el modelo DICOM-SR de informe estructurado está planteado para registrar la información de un caso, o de una serie de casos que representan el seguimiento de una patología del paciente, la HCE integra toda la información clínica del paciente a lo largo del tiempo. Esto implica información de distintas especialidades y de cómo el paciente evoluciona en cada una de ellas a lo largo de su vida. Además, todos los datos de la HCE son susceptibles de evolucionar y transformarse con el tiempo. Esta particularidad, supone un desafío para la interoperabilidad de los datos de salud entre sistemas de información sanitarios. Modelar toda la información clínica del paciente requiere de normas de HCE con modelos de referencia más sofisticados que el de DICOM-SR.

Una posible aproximación para normalizar el IRE enriquecido con la norma ISO13606, consiste en considerar el IRE como un componente COMPOSITION del modelo de referencia ISO13606 (el modelo de referencia ISO13606 se describe en detalle en el capítulo 2), de igual forma que el arquetipo de informe de resultados de pruebas de imagen del CMDIC utiliza un COMPOSITION como raíz (figura 29). Por otro lado, el único componente del MR de DICOM-

SR, el elemento de contenido, equivale a cualquiera de los siguientes componentes de ISO13606: SECTION, ENTRY, CLUSTER, ELEMENT, dependiendo de la parte de la plantilla de IRE que se esté modelando. Este abanico de posibilidades dificulta la automatización del modelado de arquetipos de HCE análogos a las plantillas de IRE (arquetipos de IRE), y la especificación de mapeo entre plantillas de IRE y arquetipos de IRE, que probablemente deba realizarse manualmente. En la literatura se encuentran trabajos en los que se plantea la conversión de objetos DICOM de secuencias nativas (no solo DICOM-SR), a instancias de HCE de acuerdo a un arquetipo [134]. Los objetos DICOM que representan parte de una secuencia nativa se modelan como un CLUSTER de ISO13606. Este enfoque permitiría añadir secuencias originales al arquetipo de IRE, que parte de un componente COMPOSITION. Una vez modelado el arquetipo de IRE, el siguiente paso es definir las relaciones que mapean la plantilla de IRE y el arquetipo de IRE. Para ello, se puede emplear la herramienta LinkEHR [19] y la representación canónica de DICOM-SR en XML [136]. Una versión del IRE enriquecido normalizado con ISO13606, como la que podría obtenerse con este método, permitiría investigar y explotar las sinergias entre la HCE aumentada con datos LOD y el IRE enriquecido con resultados de SADC.

En esta tesis se presenta un método de IRE enriquecido con resultados de SADC que se ha implantado en la práctica clínica, demostrando su viabilidad para el informado radiológico asistido con resultados de SADC. Además, se plantea cómo el IRE enriquecido puede explotarse conjuntamente con la información de salud de la HCE y de fuentes de conocimiento externas basadas en LOD. Aunque la solución de HCE aumentada presentada en esta tesis supone un paso más hacia la interoperabilidad semántica de la información de salud con otras fuentes de conocimiento potencialmente relevantes para la práctica clínica, este tipo de soluciones basadas en LOD y estándares de HCE no están ampliamente extendidas en la práctica clínica. Sin embargo, es probable que lo estén en un futuro. Esto dependerá, en gran medida, del grado de calidad de los datos disponibles en repositorios de LOD, así como del mantenimiento de los mismos en el contexto de un conocimiento cambiante y en continua evolución. Una de las principales ventajas de la solución presentada, es la capacidad de adaptarse a dicha evolución. Por ejemplo, tomando como referencia el aumento de HCE de interacciones farmacológicas presentado en esta tesis, si en el futuro se descubrieran nuevas interacciones farmacológicas para un fármaco dado, estas estarían disponibles directamente en la HCE aumentada del paciente que tenga dicho fármaco en su medicación activa, sin necesidad de redefinir la HCE aumentada, el aumento de HCE, o los arquetipos que la hicieron posible. Esta característica favorece la escalabilidad y sostenibilidad de la solución de HCE aumentada en el contexto de un conocimiento en continua evolución. El principal requisito para que esto se cumpla, es que la fuente de conocimiento que alberga dicha información sobre interacciones farmacológicas sea mantenida y/o actualizada a medida que el conocimiento evoluciona. Esto, a su vez, exige un coste considerable de mantenimiento a los responsables de las ontologías médicas y repositorios LOD, pues la cantidad de datos y conocimiento que modelan dichos repositorios crece de manera exponencial a medida que aumenta la evidencia clínica en las diferentes áreas de especialidad. La aplicación de mecanismos que permitan automatizar dicho mantenimiento es esencial para garantizar su viabilidad y sostenibilidad. Sin embargo, los procedimientos automatizados existentes en la actualidad para mantener dichos repositorios actualizados van en detrimento de la calidad de los datos que contienen en términos de integridad, consistencia, concisión e interrelación [124]. Nótese que, incluso en aquellos casos en los que el conocimiento evoluciona de tal manera que los arquetipos de una HCE aumentada quedan obsoletos, es posible utilizar los mecanismos de versionado y/o especialización definidos en el modelo ISO13606 para permitir a la HCE aumentada evolucionar en consonancia con el conocimiento, aprovechando así toda la capacidad de escalabilidad y sostenibilidad de los modelos duales de HCE.

Capítulo 6. Conclusiones y trabajo futuro

6.1. Conclusiones

La principal aportación de esta investigación radica en la integración de sistemas de HCE, resultados de SADC y datos LOD mediante estándares, herramientas públicas y metodologías formales, alcanzando en todos los casos un cierto grado de interoperabilidad entre los sistemas o fuentes de información involucrados. La literatura evidencia que se trata de un campo de interés en continua evolución para la comunidad científica. La interoperabilidad de los sistemas de HCE puede repercutir positivamente sobre la salud y el cuidado del paciente, especialmente en el paradigma de la medicina personalizada. Los objetivos y contribuciones planteados para esta tesis, ponen a disposición de los sistemas de HCE estrategias basadas en modelos de información, de arquetipos y de dominio (los tres pilares de la interoperabilidad semántica), para mejorar el grado de interoperabilidad en las integraciones de sistemas de HCE con SADC y datos LOD. En concreto, se han empleado modelos de información de HCE e IRE. Terminologías y ontologías médicas como SNOMED CT, DrugBank y NDFRT expresados en datos LOD como modelos de dominio y arquetipos ISO13606 y plantillas DICOM-SR. Con estos elementos se ha construido una solución, cuya principal cualidad es aumentar la HCE con resultados de SADC e información de fuentes externas. La HCE aumentada, además, está también normalizada lo cual asegura su interoperabilidad semántica.

La hipótesis de investigación de partida establece que la combinación y explotación conjunta de las tecnologías más avanzadas para los modelos de información, de dominio y SADC pueden facilitar la interoperabilidad entre los sistemas de HCE, los SADC y fuentes de datos LOD. Los resultados alcanzados confirman la hipótesis. En el capítulo 3 se demuestra la viabilidad de integrar información de fuentes externas de la web semántica en la HCE normalizada, garantizando la coherencia semántica mediante terminologías y relaciones entre ellas, y haciendo uso de herramientas de vanguardia para modelar, mapear y transformar información de HCE normalizada y datos LOD. En el capítulo 4 se integran resultados de SADC de diversa índole (sistemas de reglas, visión por computación y aprendizaje máquina) con el IRE, una parte fundamental de la HCE, que cuenta con su propio modelo de plantillas DICOM-SR. Por último, en el capítulo 5 se propone una aproximación sobre cómo transformar plantillas DICOM-SR de IRE a arquetipos ISO13606 de HCE. De esta manera, se abre la puerta a diversas líneas de investigación adicionales y estrategias de explotación de este trabajo, que pueden abordarse en el futuro.

Esta tesis plantea una serie de objetivos que responden a cada necesidad particular del propósito final de enriquecer la HCE con información de SADC y datos LOD.

Objetivo 1. Diseñar metodologías para enriquecer sistemas de HCE con información de SADC y datos enlazados abiertos.

Hemos diseñado dos metodologías para la consecución de este objetivo. La primera se plantea en el capítulo 3. En ella se aprovechan las características de los modelos de información de HCE, las herramientas de la web semántica, y las correspondencias entre terminologías para enriquecer HCE normalizadas con información de la web semántica. En la segunda, los resultados de diversos tipos de SADC son registrados en sofisticadas plantillas de IRE, que posteriormente se exportan en formato normalizado DICOM-SR.

Objetivo 1.1. Definir transformaciones para la integración de información de SADC en la HCE.

En el capítulo 4, se presenta una metodología para embeber en las plantillas de IRE enriquecido los resultados de SADC relevantes para cada tipo de plantilla. Los IRE informados con este sistema se pueden exportar en DICOM-SR, y sería posible transformarlos posteriormente al formato de HCE normalizada ISO13606 siguiendo la propuesta del capítulo 5.

Objetivo 1.2. Definir transformaciones para la integración de información de dominio en la HCE.

La metodología para obtener la HCE aumentada, presentada en el capítulo 3, establece las transformaciones o intercambios de datos necesarios para integrar información de dominio en la HCE de forma generalizada. En concreto, se definen dos transformaciones, (I) convertir instancias de datos LOD en formato de HCE normalizada (contenido de aumento normalizado), y (II) integrar el contenido de aumento normalizado junto con la instancia de HCE inicial, dando lugar a la HCE aumentada. Estas transformaciones aprovechan los arquetipos de HCE ISO13606, y el esquema canónico del W3C para resultados de consultas de SPARQL en XML, como esquemas de partida.

Objetivo 2. Desarrollar el concepto de HCE aumentada y mecanismos para la creación, uso, persistencia y consulta de la HCE aumentada.

La HCE aumentada se desarrolla a lo largo de varios capítulos de este trabajo, aunque su definición conceptual y metodológica se describe en el capítulo 3. En ese capítulo se plantean métodos para su creación, uso, persistencia y consulta, y se proporciona un caso de uso de HCE aumentada con diferentes aumentos de HCE.

Objetivo 2.1. Definir conceptualmente la HCE aumentada.

La HCE aumentada (o enriquecida), entendida como una HCE existente complementada con contenidos adicionales potencialmente relevantes y procedentes de fuentes de información externas, se define detalladamente en el capítulo 3. La metodología parte de un arquetipo de HCE y se basa en la identificación y definición formal de enlaces y transformaciones entre una HCE y una fuente de información de la web semántica.

Objetivo 2.2. Aumentar la HCE con datos enlazados abiertos.

En el capítulo 3, ponemos en práctica el método para obtener la HCE aumentada mediante un caso de uso. La HCE resumida del CMDIC, es aumentada con información sobre interacciones farmacológicas para la medicación activa del paciente, e información de posibles tratamientos para los episodios activos del paciente. En este caso de uso, la información adicional se obtiene mediante representaciones LOD de las terminologías y ontologías SNOMED CT, DrugBank y NDFRT.

Objetivo 2.3. Proporcionar mecanismos de persistencia y consulta de la HCE aumentada.

Se han empleado sistemas de persistencia basados en XML para el almacenamiento, consulta, modificación y borrado de aumentos de HCE. La elección de XML para la capa de persistencia ha sido motivada por el formato de las diferentes partes de un aumento de HCE, conformado principalmente por scripts XQuery y arquetipos definidos en lenguaje ADL. Tanto los scripts XQuery como los arquetipos expresados en ADL son documentos que pueden embeberse fácilmente en un extracto XML. Una capa de persistencia basada en XML favorece la búsqueda e indexación de aumentos de HCE tanto por su identificador como por su contenido. El lenguaje XQuery permite interactuar con esta base de datos, y el caso de uso implementado (capítulo 3) utiliza XQuery para consultar aumentos de HCE. Por ejemplo, identificar los aumentos disponibles para un determinado extracto de HCE, o la versión más reciente de un aumento de HCE.

Objetivo 3. Desarrollar el concepto de IRE enriquecido con resultados de SADC y mecanismos para la creación, uso, persistencia y consulta del IRE enriquecido.

El concepto de IRE enriquecido se desarrolla a lo largo de varios capítulos de esta tesis, siendo en el capítulo 4 donde se describe detalladamente su definición y alcance. Además, se presenta un sistema de IRE enriquecido con resultados de SADC, que proporciona mecanismos para la creación, uso, persistencia y consulta del IRE enriquecido. La implantación de este sistema en 37 instituciones (enumeradas en el Anexo II) para el informado radiológico demuestra su viabilidad.

Objetivo 3.1. Definir conceptualmente el IRE enriquecido.

En el capítulo 4, se define el IRE enriquecido como un IRE convencional que integra resultados de diversos tipos de SADC: basados en reglas, en técnicas de visión por computación y en modelos de aprendizaje automático. Además, nuestro concepto de IRE enriquecido contempla mecanismos para la validación de datos introducidos por el usuario, automatiza la obtención de valores calculables a partir de dichos datos, y permite generar una versión del contenido del informe expresada en lenguaje natural. Estas características sirven para guiar el informado, contribuir a minimizar errores, y adaptar el contenido del informe a la audiencia objetivo.

Objetivo 3.2. Enriquecer el IRE con resultados de SADC.

La solución presentada en el capítulo 4, permite integrar en el IRE resultados de SADC basados en cálculos prescritos en guías radiológicas, datos cuantitativos de biomarcadores de imagen, e inferencia de potenciales lesiones con modelos de aprendizaje automático. En concreto, se describe un caso de uso con el IRE para la evaluación del cáncer de próstata con IRM. Este es enriquecido con: (I) un sistema semi-automático de cálculo de puntuación PI-RADS 2.1, (II) datos cuantitativos de la fracción de v_e , ADC, K^{trans} , k_{ep} , y (III) potenciales lesiones en la próstata inferidas mediante redes neuronales convolucionales.

Objetivo 3.3. Proporcionar mecanismos de persistencia y consulta del IRE enriquecido.

El diseño del sistema de IRE enriquecido se plantea de forma modular (capítulo 4), facilitando la creación de plantillas de IRE enriquecidas y personalizadas. Nuestra solución gestiona la creación, especialización, versionado y borrado de plantillas de IRE de acuerdo al perfil de integración de IHE MRRT. Además, utiliza una base de datos MongoDB para persistir, recuperar, consultar o borrar las plantillas de IRE. Por último, los IRE se pueden exportar en formato DICOM-SR o en formato de mensaje HL7 v2. Esto facilita la integración del IRE enriquecido con otros sistemas de información sanitarios que implementen las especificaciones DICOM y/o HL7 v2.

6.2. Trabajo futuro

El diseño de una HCE aumentada requiere de ciertas tareas de modelado manuales que impiden su automatización completa. No obstante, existen herramientas que pueden asistir al proceso de diseño de HCE aumentada; como el uso del rastreador de datos enlazados LDSpider [137] para descubrir contenidos de aumento potencialmente interesantes a partir de la EDI. Además, los editores visuales de consultas SPARQL, como Gruff [138], son útiles

para explorar el contenido de aumento y construir las consultas SPARQL adecuadas. Los métodos de IA son otra línea de investigación futura que puede contribuir a automatizar la definición de aumentos de HCE o a reducir las intervenciones manuales. La investigación futura debe centrarse en definir nuevos enfoques para automatizar el diseño de HCE aumentadas; por ejemplo, herramientas para explorar automáticamente las anotaciones terminológicas de un esquema de HCE, y utilizarlas para descubrir información potencialmente relevante en la LODC. Esta aproximación, además, puede servir para determinar qué aumentos de HCE, de entre los disponibles, son adecuados para enriquecer un extracto de HCE dado.

Como trabajo futuro, puede ser interesante investigar y evaluar el rendimiento del sistema de IRE enriquecido con resultados de SADC en la práctica clínica. Midiendo factores como la satisfacción de los usuarios o el tiempo necesario para el informado, en comparación con los sistemas convencionales de dictado. Asimismo, es de potencial interés investigar cómo el IRE enriquecido fomenta la radiómica, en contraste con la elaboración de informes en lenguaje natural, aprovechando las ventajas de los datos estructurados y los resultados de SADC. Estudios de mercado adicionales para comparar el sistema de IRE enriquecido con soluciones comerciales existentes, pueden ayudar a evaluar el impacto real de este trabajo. Sin embargo, en el momento de escribir esta tesis, en la literatura no se han identificado alternativas a nuestro sistema de IRE enriquecido que proporcionen las mismas funcionalidades juntas.

La normalización del IRE en formato de HCE estándar es una línea futura de investigación, continuista con el trabajo presentado en esta tesis, que implica la formalización del método propuesto en el capítulo 5 para normalizar el IRE con la norma ISO13606, así como los desarrollos necesarios para su prueba de concepto y validación. Por último, otra línea futura de investigación es la creación de un repositorio de datos clínicos estructurados, procedentes de fuentes de datos multidisciplinares, y obtenidos aplicando el sistema de informe estructurado en otros dominios clínicos distintos de la radiología (como los informes en patología o genética). Un repositorio de estas características sería potencialmente relevante para la minería de datos, el entrenamiento de SADC basados en aprendizaje automático, o la explotación del método de HCE aumentada para obtener contenido de aumento de la web semántica relacionado con cualquier dato (o combinación de datos) de dicho repositorio.

Bibliografía

- [1] D. de Boer, D. Delnoij, y J. Rademakers, «The importance of patient-centered care for various patient groups», *Patient Educ. Couns.*, vol. 90, n.º 3, pp. 405-410, 2013, doi: 10.1016/j.pec.2011.10.002.
- [2] S. Bowman, «Impact of electronic health record systems on information integrity: quality and safety implications.», *Perspect. Health Inf. Manag.*, vol. 10, n.º Fall, pp. 1c-1c, 2013.
- [3] A. Mañas-García, J. A. Maldonado, M. Marcos, D. Boscá, y M. Robles, «Augmented EHR: Enrichment of EHR with Contents from Semantic Web Sources», *Appl. Sci.*, vol. 11, n.º 9, Art. n.º 9, ene. 2021, doi: 10.3390/app11093978.
- [4] D. Boscá, J. A. Maldonado, D. Moner, y M. Robles, «Automatic generation of computable implementation guides from clinical information models», *J. Biomed. Inform.*, vol. 55, pp. 143-152, jun. 2015, doi: 10.1016/j.jbi.2015.04.002.
- [5] M. Marcos, J. A. Maldonado, B. Martínez-Salvador, D. Moner, D. Boscá, y M. Robles, «An Archetype-Based Solution for the Interoperability of Computerised Guidelines and Electronic Health Records», *Artif. Intell. Med. Vol. 6747 Ser. Lect. Notes Comput. Sci. Pp 276-285*, pp. 276-285, 2011, doi: 10.1007/978-3-642-22218-4_35.
- [6] M. Marcos, J. A. Maldonado, B. Martínez-Salvador, D. Boscá, y M. Robles, «Interoperability of clinical decision-support systems and electronic health records using archetypes: a case study in clinical trial eligibility.», *J. Biomed. Inform.*, vol. 46, n.º 4, pp. 676-89, ago. 2013, doi: 10.1016/j.jbi.2013.05.004.
- [7] A. Seyfang, B. Martínez-Salvador, R. Serban, J. Wittenberg, S. Miksch, M. Marcos, A. ten Teije, y K. Rosenbrand, «Maintaining Formal Models of Living Guidelines Efficiently», en *Artificial Intelligence in Medicine*, Berlin, Heidelberg, 2007, pp. 441-445. doi: 10.1007/978-3-540-73599-1_58.
- [8] A. Mañas-García, I. González-Valverde, E. Camacho-Ramos, A. Alberich-Bayarri, J. A. Maldonado, M. Marcos, y M. Robles, «Radiological Structured Report Integrated with Quantitative Imaging Biomarkers and Qualitative Scoring Systems», *J. Digit. Imaging*, feb. 2022, doi: 10.1007/s10278-022-00589-9.
- [9] A. T. Teije, S. Miksch, y P. Lucas, *Computer-based Medical Guidelines and Protocols: A Primer and Current Trends*. IOS Press, 2008.
- [10] A. Mañas-García, E. Camacho-Ramos, I. Gonzalez-Valverde, F. Bacha-Villamide, A. Alberich-Bayarri, y L. Martí-Bonmatí, «Implementation of an interactive radiological structured report management system with AI annotation capabilities», presentado en European Society of Medical Imaging Informatics (EuSoMII) Annual Meeting 2019. Medical Imaging Informatics. AI, Clinical Applications and more., 2019.
- [11] B. Martínez-Salvador, M. Marcos, A. Mañas-García, J. A. Maldonado, y M. Robles,

- «Using SNOMED CT Expression Constraints to Bridge the Gap Between Clinical Decision-Support Systems and Electronic Health Records.», *Stud. Health Technol. Inform.*, vol. 228, pp. 504-8, 2016.
- [12] A. Mañas-García, R. S. Requena, A. Alberich-Bayarri, G. García-Martí, L. A. Tortajada, E. R. Martínez, y L. Martí-Bonmatí, «Imaging biomarkers processing platform», en *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, jun. 2014, pp. 416-419. doi: 10.1109/BHI.2014.6864391.
- [13] R. Sanz-Requena, A. Mañas-García, M. Robles Viejo, y L. Martí-Bonmatí, «Web application for the integration of bone age calculation in RIS/PACS», *ECR 2013 EPOS*, 7 de marzo de 2013. <https://epos.myesr.org/poster/esr/ecr2013/C-0493> (accedido 20 de febrero de 2022).
- [14] A. Mañas-García, R. Sanz-Requena, y M. Robles Viejo, «Aplicación Web para la integración del cálculo de la edad ósea con los sistemas de información del entorno hospitalario», presentado en XXX Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2012), San Sebastián, España, 2012.
- [15] C. Martínez-Costa, M. Menárguez-Tortosa, J. T. Fernández-Breis, y J. A. Maldonado, «A model-driven approach for representing clinical archetypes for Semantic Web environments», *J. Biomed. Inform.*, vol. 42, n.º 1, pp. 150-164, feb. 2009, doi: 10.1016/j.jbi.2008.05.005.
- [16] J. T. Fernández-Breis, J. A. Maldonado, M. Marcos, M. del C. Legaz-García, D. Moner, J. Torres-Sospedra, A. Esteban-Gil, B. Martínez-Salvador, y M. Robles, «Leveraging electronic healthcare record standards and semantic web technologies for the identification of patient cohorts.», *J. Am. Med. Inform. Assoc. JAMIA*, vol. 20, n.º e2, pp. e288-96, dic. 2013, doi: 10.1136/amiajnl-2013-001923.
- [17] J. A. Maldonado, D. Moner, D. Boscá, J. T. Fernández-Breis, C. Angulo, y M. Robles, «LinkEHR-Ed: A multi-reference model archetype editor based on formal semantics», *Int. J. Med. Inf.*, vol. 78, n.º 8, pp. 559-570, ago. 2009, doi: 10.1016/j.ijmedinf.2009.03.006.
- [18] L. Marco-Ruiz, D. Moner, J. A. Maldonado, N. Kolstrup, y J. G. Bellika, «Archetype-based data warehouse environment to enable the reuse of electronic health record data.», *Int. J. Med. Inf.*, vol. 84, n.º 9, pp. 702-14, sep. 2015, doi: 10.1016/j.ijmedinf.2015.05.016.
- [19] J. A. Maldonado, D. Moner, D. Bosca, C. Angulo, L. Marco, E. Reig, y M. Robles, «Concept-Based Exchange of Healthcare Information: The LinkEHR Approach», en *2011 IEEE First International Conference on Healthcare Informatics, Imaging and Systems Biology*, jul. 2011, pp. 150-157. doi: 10.1109/HISB.2011.18.
- [20] J. A. Maldonado, C. M. Costa, D. Moner, M. Menárguez-Tortosa, D. Boscá, J. A.

- Miñarro Giménez, J. T. Fernández-Breis, y M. Robles, «Using the ResearchEHR platform to facilitate the practical application of the EHR standards», *J. Biomed. Inform.*, vol. 45, n.º 4, pp. 746-762, ago. 2012, doi: 10.1016/j.jbi.2011.11.004.
- [21] A. ten Teije, M. Marcos, M. Balsler, *et al.*, «Improving medical protocols by formal methods», *Artif. Intell. Med.*, vol. 36, n.º 3, pp. 193-209, mar. 2006, doi: 10.1016/j.artmed.2005.10.006.
- [22] J. C. Crespo, A. Mañas-García, D. B. Tomás, y M. Robles Viejo, «Metodología para el modelado de Recursos HL7 FHIR mediante arquetipos del estándar ISO13606», en *CASEIB 2016 XXXIV Congreso Anual de la Sociedad Española de Ingeniería Biomédica: Libro de actas, 2016, ISBN 978-84-9048-531-6, págs. 147-150*, 2016, pp. 147-150. Accedido: 20 de febrero de 2022. [En línea]. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=8271615>
- [23] B. H. de Mello, S. J. Rigo, C. A. da Costa, R. da Rosa Righi, B. Donida, M. R. Bez, y L. C. Schunke, «Semantic interoperability in health records standards: a systematic literature review», *Health Technol.*, ene. 2022, doi: 10.1007/s12553-022-00639-w.
- [24] «SNOMED International». [En línea]. Disponible en: <http://www.snomed.org/snomed-ct>
- [25] «DrugBank. A pharmaceutical knowledge base». [En línea]. Disponible en: <https://www.drugbank.ca/>
- [26] «National Drug File - Reference Terminology Source Information». [En línea]. Disponible en: <https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/NDFRT/index.html>
- [27] D. L. Beaty y D. Quirk, «The digital revolution», *ASHRAE J.*, vol. 57, n.º 5, pp. 80-85, may 2015.
- [28] M. Barros y F. M. Couto, «Knowledge Representation and Management: a Linked Data Perspective», *Yearb. Med. Inform.*, vol. 25, n.º 1, pp. 178-183, ago. 2016, doi: 10.15265/IY-2016-022.
- [29] I. E. de S. y B. Social, «Sistemas de información sanitaria: caminando hacia el futuro |», *Instituto Europeo de Salud y Bienestar Social*, 20 de enero de 2021. <https://institutoeuropeo.es/articulos/sistemas-de-informacion-sanitaria-caminando-hacia-el-futuro/> (accedido 1 de mayo de 2022).
- [30] K. S. Mann y A. Bansal, «HIS Integration Systems Using Modality Worklist and DICOM», *Procedia Comput. Sci.*, vol. 37, pp. 16-23, ene. 2014, doi: 10.1016/j.procs.2014.08.007.
- [31] B. Blazona y M. Koncar, «HL7 and DICOM based integration of radiology departments with healthcare enterprise information systems», *Int. J. Med. Inf.*, vol. 76, pp. S425-S432, dic. 2007, doi: 10.1016/j.ijmedinf.2007.05.001.
- [32] ISO, «Technical Report 20514: Electronic health record: Definition, scope and context»,

- ene. 2005.
- [33] «The CEN/ISO EN13606 standard». [En línea]. Disponible en: <http://www.en13606.org/the-ceniso-en13606-standard>
- [34] I. Landi, J. De Freitas, B. A. Kidd, J. T. Dudley, B. S. Glicksberg, y R. Miotto, «The Evolution of Mining Electronic Health Records in the Era of Deep Learning», en *Deep Learning in Biology and Medicine*, WORLD SCIENTIFIC (EUROPE), 2021, pp. 55-92. doi: 10.1142/9781800610941_0003.
- [35] A. Gutiérrez, G. Cuenca, I. Acebedo, J. Tello, R. Madariaga, P. Serrano-Balazote, D. Moner, y J. Maldonado, *Manual práctico de interoperabilidad semántica para entornos sanitarios basados en arquetipos*. 2013, p. 101.
- [36] A. Moreno-Conde, D. Moner, W. D. da Cruz, M. R. Santos, J. A. Maldonado, M. Robles, y D. Kalra, «Clinical information modeling processes for semantic interoperability of electronic health records: systematic review and inductive analysis.», *J. Am. Med. Inform. Assoc. JAMIA*, vol. 22, n.º 4, pp. 925-34, jul. 2015, doi: 10.1093/jamia/ocv008.
- [37] Ministerio de Sanidad y Política Social, *Real Decreto 1093/2010, de 3 de septiembre, por el que se aprueba el conjunto mínimo de datos de los informes clínicos en el Sistema Nacional de Salud*, vol. BOE-A-2010-14199. 2010, pp. 78742-78767. Accedido: 21 de marzo de 2022. [En línea]. Disponible en: <https://www.boe.es/eli/es/rd/2010/09/03/1093>
- [38] «openEHR Home». <https://www.openehr.org/> (accedido 17 de noviembre de 2021).
- [39] «HL7 Standards Product Brief - CDA® Release 2 | HL7 International». http://www.hl7.org/implement/standards/product_brief.cfm?product_id=7 (accedido 7 de abril de 2022).
- [40] «Health Level Seven International - FHIR v4.0.1». Accedido: 7 de abril de 2022. [En línea]. Disponible en: <https://www.hl7.org/fhir/>
- [41] NEMA, «Digital Imaging and Communication in Medicine», *DICOM Standard*. <https://www.dicomstandard.org/current> (accedido 30 de marzo de 2022).
- [42] W. D. Bidgood, S. C. Horii, F. W. Prior, y D. E. Van Syckle, «Understanding and Using DICOM, the Data Interchange Standard for Biomedical Imaging», *J. Am. Med. Inform. Assoc.*, vol. 4, n.º 3, pp. 199-212, 1997.
- [43] A. Krishnaraj, J. C. Weinreb, P. H. Ellenbogen, B. Allen, A. Norbash, y E. A. Kazerooni, «The future of imaging biomarkers in radiologic practice: Proceedings of the thirteenth Annual ACR forum», *J. Am. Coll. Radiol.*, vol. 11, n.º 1, pp. 20-23, ene. 2014, doi: 10.1016/j.jacr.2013.08.017.
- [44] J. Y. An, K. M. L. Unsdorfer, y J. C. Weinreb, «BI-RADS, C-RADS, CAD-RADS, LI-RADS, Lung-RADS, NI-RADS, O-RADS, PI-RADS, TI-RADS: Reporting and Data Systems», *RadioGraphics*, vol. 39, n.º 5, pp. 1435-1436, sep. 2019, doi:

- 10.1148/rg.2019190087.
- [45] J. Kemesiene y C. Nicolau, «Prostate MRI for dummies», *ECR 2019 EPOS*, 11 de enero de 2019. <https://epos.myesr.org/poster/esr/ecr2019/C-3320> (accedido 31 de marzo de 2022).
- [46] M. Kasel-Seibert, T. Lehmann, R. Aschenbach, F. V. Guettler, M. Abubrig, M.-O. Grimm, U. Teichgraeber, y T. Franiel, «Assessment of PI-RADS v2 for the Detection of Prostate Cancer», *Eur. J. Radiol.*, vol. 85, n.º 4, pp. 726-731, abr. 2016, doi: 10.1016/j.ejrad.2016.01.011.
- [47] T. Pun, G. Gerig, y O. Ratib, «Image analysis and computer vision in medicine», *Comput. Med. Imaging Graph.*, vol. 18, n.º 2, pp. 85-96, mar. 1994, doi: 10.1016/0895-6111(94)90017-5.
- [48] M. A. Jacobs, R. Ouwerkerk, K. Petrowski, y K. J. Macura, «Diffusion Weighted Imaging with ADC Mapping and Spectroscopy in Prostate Cancer», *Top. Magn. Reson. Imaging TMRI*, vol. 19, n.º 6, pp. 261-272, dic. 2008, doi: 10.1097/RMR.0b013e3181aa6b50.
- [49] P. Pelechano-Gómez, «Optimización de la secuencia de difusión en RM 3T con modelo multifactorial IVIM en el estudio de la próstata.», Universitat de València, 2017. [En línea]. Disponible en: <https://roderic.uv.es/bitstream/handle/10550/59087/TESIS%20DOCTORAL.%20PAULA%20PELECHANO.pdf?sequence=1>
- [50] I. Ocak, M. Bernardo, G. Metzger, T. Barrett, P. Pinto, P. S. Albert, y P. L. Choyke, «Dynamic Contrast-Enhanced MRI of Prostate Cancer at 3 T: A Study of Pharmacokinetic Parameters», *Am. J. Roentgenol.*, vol. 189, n.º 4, pp. W192-W201, oct. 2007, doi: 10.2214/AJR.06.1329.
- [51] O. Chilali, P. Puech, S. Lakroum, M. Diaf, S. Mordon, y N. Betrouni, «Gland and Zonal Segmentation of Prostate on T2W MR Images», *J. Digit. Imaging*, vol. 29, n.º 6, pp. 730-736, dic. 2016, doi: 10.1007/s10278-016-9890-0.
- [52] N. A. Obuchowski, A. P. Reeves, E. P. Huang, *et al.*, «Quantitative imaging biomarkers: A review of statistical methods for computer algorithm comparisons», *Stat. Methods Med. Res.*, vol. 24, n.º 1, pp. 68-106, feb. 2015, doi: 10.1177/0962280214537390.
- [53] «IA para interpretar imágenes médicas • bit», *bit*, 31 de marzo de 2020. <https://bit.coit.es/ia-para-interpretar-imagenes-medicas/> (accedido 1 de abril de 2022).
- [54] M. L. Giger, «Machine Learning in Medical Imaging», *J. Am. Coll. Radiol.*, vol. 15, n.º 3, Part B, pp. 512-520, mar. 2018, doi: 10.1016/j.jacr.2017.12.028.
- [55] D. Ganeshan, P. A. T. Duong, L. Probyn, *et al.*, «Structured Reporting in Radiology», *Acad. Radiol.*, vol. 25, n.º 1, pp. 66-73, ene. 2018, doi: 10.1016/j.acra.2017.08.005.
- [56] J. M. L. Bosmans, E. Neri, O. Ratib, y C. E. Kahn, «Structured reporting: a fusion reactor hungry for fuel», *Insights Imaging*, vol. 6, n.º 1, pp. 129-132, feb. 2015, doi:

- 10.1007/s13244-014-0368-7.
- [57] L. H. Schwartz, D. M. Panicek, A. R. Berk, Y. Li, y H. Hricak, «Improving communication of diagnostic radiology findings through structured reporting», *Radiology*, vol. 260, n.º 1, pp. 174-181, jul. 2011, doi: 10.1148/radiol.11101913.
- [58] «DICOM Structured Reporting». <http://www.pixelmed.com/srbook.html> (accedido 26 de octubre de 2021).
- [59] T. Martin-Carreras, T. S. Cook, y C. E. Kahn, «Readability of radiology reports: implications for patient-centered care», *Clin. Imaging*, vol. 54, pp. 116-120, mar. 2019, doi: 10.1016/j.clinimag.2018.12.006.
- [60] C. E. Kahn, M. E. Heilbrun, y K. E. Applegate, «From guidelines to practice: How reporting templates promote the use of radiology practice guidelines», *J. Am. Coll. Radiol.*, vol. 10, n.º 4, pp. 268-273, abr. 2013, doi: 10.1016/j.jacr.2012.09.025.
- [61] C. M. Hawkins, S. Hall, B. Zhang, y A. J. Towbin, «Creation and Implementation of Department-Wide Structured Reports: An Analysis of the Impact on Error Rate in Radiology Reports», *J. Digit. Imaging*, vol. 27, n.º 5, pp. 581-587, may 2014, doi: 10.1007/s10278-014-9699-7.
- [62] D. B. Larson, A. J. Towbin, R. M. Pryor, y L. F. Donnelly, «Improving consistency in radiology reporting through the use of department-wide standardized structured reporting», *Radiology*, vol. 267, n.º 1, pp. 240-250, abr. 2013, doi: 10.1148/radiol.12121502.
- [63] N. Manoonchai, R. Kaewlai, A. Wibulpolprasert, U. Boonpramarn, A. Tohmee, y S. Phongkitkarun, «Satisfaction of Imaging Report Rendered in Emergency Setting: A Survey of Radiology and Referring Physicians», *Acad. Radiol.*, vol. 22, n.º 6, pp. 760-770, jun. 2015, doi: 10.1016/j.acra.2015.01.006.
- [64] D. Pinto dos Santos, J.-M. Hempel, P. Mildenerger, R. Klöckner, y T. Persigehl, «Structured Reporting in Clinical Routine», *RöFo - Fortschritte Auf Dem Geb. Röntgenstrahlen Bildgeb. Verfahr.*, vol. 191, n.º 01, pp. 33-39, ene. 2019, doi: 10.1055/a-0636-3851.
- [65] A. J. Towbin y C. M. Hawkins, «Use of a Web-Based Calculator and a Structured Report Generator to Improve Efficiency, Accuracy, and Consistency of Radiology Reporting», *J. Digit. Imaging*, vol. 30, n.º 5, pp. 584-588, oct. 2017, doi: 10.1007/s10278-017-9967-4.
- [66] L. Martí-Bonmatí, E. Ruiz-Martínez, A. Ten, y A. Alberich-Bayarri, «How to integrate quantitative information into imaging reports for oncologic patients», *Radiologia*, vol. 60, pp. 43-52, 2018, doi: 10.1016/j.rx.2018.02.005.
- [67] T. K. Alkasab, B. C. Bizzo, L. L. Berland, S. Nair, P. V. Pandharipande, y H. B. Harvey, «Creation of an Open Framework for Point-of-Care Computer-Assisted Reporting and

- Decision Support Tools for Radiologists», *J. Am. Coll. Radiol.*, vol. 14, n.º 9, pp. 1184-1189, sep. 2017, doi: 10.1016/j.jacr.2017.04.031.
- [68] J. Y. Chen, T. M. Sippel Schmidt, C. D. Carr, y C. E. Kahn, «Enabling the Next-Generation Radiology Report: Description of Two New System Standards», *RadioGraphics*, vol. 37, n.º 7, pp. 2106-2112, nov. 2017, doi: 10.1148/rg.2017160106.
- [69] C. E. Kahn, B. Genereaux, y C. P. Langlotz, «Conversion of Radiology Reporting Templates to the MRRT Standard», *J. Digit. Imaging*, vol. 28, n.º 5, pp. 528-536, oct. 2015, doi: 10.1007/s10278-015-9787-3.
- [70] D. Pinto dos Santos, G. Klos, R. Kloeckner, R. Oberle, C. Dueber, y P. Mildenberger, «Development of an IHE MRRT-compliant open-source web-based reporting platform», *Eur. Radiol.*, vol. 27, n.º 1, pp. 424-430, ene. 2017, doi: 10.1007/s00330-016-4344-0.
- [71] D. Pinto dos Santos, S. Scheibl, G. Arnhold, A. Maehringer-Kunz, C. Düber, P. Mildenberger, y R. Kloeckner, «A proof of concept for epidemiological research using structured reporting with pulmonary embolism as a use case», *Br. J. Radiol.*, vol. 91, n.º 1088, p. bjr.20170564-bjr.20170564, jun. 2018, doi: 10.1259/bjr.20170564.
- [72] «C.17.3 SR Document Content Module». https://dicom.nema.org/medical/Dicom/2018d/output/chtml/part03/sect_C.17.3.html (accedido 6 de febrero de 2022).
- [73] «A Structured Reporting Templates (Normative)». https://dicom.nema.org/medical/dicom/current/output/chtml/part16/chapter_A.html (accedido 7 de abril de 2022).
- [74] RSNA, «RadReport Template Library. Structured templates for clear and consistent reports.» [En línea]. Disponible en: <https://radreport.org/>
- [75] G. W. Boland, D. R. Enzmann, y R. Duszak, «Actionable reporting», *J. Am. Coll. Radiol. JACR*, vol. 11, n.º 9, pp. 844-845, sep. 2014, doi: 10.1016/j.jacr.2014.06.002.
- [76] G. W. Boland y R. Duszak, «Structured reporting and communication», *J. Am. Coll. Radiol.*, vol. 12, n.º 12, pp. 1286-1288, dic. 2015, doi: 10.1016/j.jacr.2015.08.001.
- [77] A. Pomar-Nadal, C. Pérez-Castillo, A. Alberich-Bayarri, G. García-Martí, R. Sanz Requena, y L. Martí-Bonmatí, «Integrating information about imaging biomarkers into structured radiology reports», *Radiologia*, vol. 55, n.º 3, pp. 188-194, may 2013, doi: 10.1016/j.rx.2012.11.005.
- [78] «Healthcare Information and Management Systems Society | HIMSS». <https://www.himss.org/> (accedido 25 de octubre de 2021).
- [79] HIMSS, «Definition of Interoperability Approved by the HIMSS Board of Directors», April 5, 2013. [En línea]. Disponible en: [http://www.himss.org/sites/himssorg/files/FileDownloads/HIMSS Interoperability Definition FINAL.pdf](http://www.himss.org/sites/himssorg/files/FileDownloads/HIMSS%20Interoperability%20Definition%20FINAL.pdf)

- [80] Directorate-General for the Information Society and Media (European Commission) Now known as, M. Virtanen, B. Ustun, *et al.*, *Semantic interoperability for better health and safer healthcare: deployment and research roadmap for Europe*. LU: Publications Office of the European Union, 2009. Accedido: 25 de octubre de 2021. [En línea]. Disponible en: <https://data.europa.eu/doi/10.2759/38514>
- [81] A. Wright, D. F. Sittig, J. S. Ash, *et al.*, «Lessons learned from implementing service-oriented clinical decision support at four sites: A qualitative study», *Int. J. Med. Inf.*, vol. 84, n.º 11, pp. 901-911, 2015, doi: 10.1016/j.ijmedinf.2015.08.008.
- [82] P. Muñoz, J. Trigo, I. Martínez, A. Muñoz, J. Escayola, y J. García, «The ISO/EN 13606 Standard for the Interoperable Exchange of Electronic Health Records», *J. Healthc. Eng.*, vol. 2, n.º 1, pp. 1-24, mar. 2011, doi: 10.1260/2040-2295.2.1.1.
- [83] L. Marco-Ruiz, C. Pedrinaci, J. A. Maldonado, L. Panziera, R. Chen, y J. G. Bellika, «Publication, discovery and interoperability of Clinical Decision Support Systems: A Linked Data approach», *J. Biomed. Inform.*, vol. 62, pp. 243-264, 2016, doi: 10.1016/j.jbi.2016.07.011.
- [84] M. Marcos y B. Martínez-Salvador, «Towards the interoperability of computerised guidelines and electronic health records: An experiment with openEHR archetypes and a chronic heart failure guideline», *Lect. Notes Comput. Sci. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinforma.*, vol. 6512 LNAI, pp. 101-113, 2011, doi: 10.1007/978-3-642-18050-7_8.
- [85] eHealth Network, «Refined eHealth European Interoperability Framework». 2015. Accedido: 26 de octubre de 2021. [En línea]. Disponible en: https://ec.europa.eu/health/sites/default/files/ehealth/docs/ev_20151123_co03_en.pdf
- [86] M. Herrero-Zazo, I. Segura-Bedmar, J. Hastings, y P. Martínez, «DINTO: Using OWL Ontologies and SWRL Rules to Infer Drug-Drug Interactions and Their Mechanisms.», *J. Chem. Inf. Model.*, vol. 55, n.º 8, pp. 1698-707, ago. 2015, doi: 10.1021/acs.jcim.5b00119.
- [87] J. M. Banda, A. Callahan, R. Winnenburg, *et al.*, «Feasibility of Prioritizing Drug-Drug-Event Associations Found in Electronic Health Records.», *Drug Saf.*, vol. 39, n.º 1, pp. 45-57, ene. 2016, doi: 10.1007/s40264-015-0352-2.
- [88] C. M. Costa, M. Menárguez-Tortosa, y J. T. Fernández-Breis, «Clinical data interoperability based on archetype transformation.», *J. Biomed. Inform.*, vol. 44, n.º 5, pp. 869-80, oct. 2011, doi: 10.1016/j.jbi.2011.05.006.
- [89] L. Lezcano, M.-A. Sicilia, y C. Rodríguez-Solano, «Integrating reasoning and clinical archetypes using OWL ontologies and SWRL rules.», *J. Biomed. Inform.*, vol. 44, n.º 2, pp. 343-53, abr. 2011, doi: 10.1016/j.jbi.2010.11.005.
- [90] B. Haarbrandt, T. Jack, y M. Marschollek, «Automated Transformation of openEHR

- Data Instances to OWL.», *Stud. Health Technol. Inform.*, vol. 223, pp. 63-70, ene. 2016.
- [91] C. Tao, G. Jiang, T. A. Oniki, R. R. Freimuth, Q. Zhu, D. Sharma, J. Pathak, S. M. Huff, y C. G. Chute, «A semantic-web oriented representation of the clinical element model for secondary use of electronic health records data.», *J. Am. Med. Inform. Assoc. JAMIA*, vol. 20, n.º 3, pp. 554-62, may 2013, doi: 10.1136/amiajnl-2012-001326.
- [92] J. Pathak, R. C. Kiefer, y C. G. Chute, «Applying linked data principles to represent patient's electronic health records at Mayo clinic», en *Proceedings of the 2nd ACM SIGHIT symposium on International health informatics - IHI '12*, New York, New York, USA, 2012, pp. 455-455. doi: 10.1145/2110363.2110415.
- [93] M. del Carmen Legaz-García, J. A. Miñarro-Giménez, M. Menárguez-Tortosa, *et al.*, «Generation of open biomedical datasets through ontology-driven transformation and integration processes», *J. Biomed. Semant.*, vol. 7, n.º 1, pp. 32-32, dic. 2016, doi: 10.1186/s13326-016-0075-z.
- [94] B. Koopman, P. Bruza, L. Sitbon, y M. Lawley, «Towards semantic search and inference in electronic medical records: An approach using concept--based information retrieval.», *Australas. Med. J.*, vol. 5, n.º 9, pp. 482-8, ene. 2012, doi: 10.4066/AMJ.2012.1362.
- [95] J. L. Allones, M. Taboada, D. Martinez, R. Lozano, y M. J. Sobrido, «SNOMED CT module-driven clinical archetype management.», *J. Biomed. Inform.*, vol. 46, n.º 3, pp. 388-400, jun. 2013, doi: 10.1016/j.jbi.2013.01.003.
- [96] M. Menárguez-Tortosa y J. T. Fernández-Breis, «OWL-based reasoning methods for validating archetypes.», *J. Biomed. Inform.*, vol. 46, n.º 2, pp. 304-17, abr. 2013, doi: 10.1016/j.jbi.2012.11.009.
- [97] A. M. Porn, L. M. Peres, y M. Didonet Del Fabro, «A Process for the Representation of openEHR ADL Archetypes in OWL Ontologies», *Studies in health technology and informatics*, vol. 216, pp. 827-31, enero de 2015. [En línea]. Disponible en: <http://ebooks.iospress.nl/publication/40325>
- [98] D. J. Odgers y M. Dumontier, «Mining Electronic Health Records using Linked Data.», *AMIA Jt. Summits Transl. Sci. Proc. AMIA Summit Transl. Sci.*, vol. 2015, pp. 217-21, ene. 2015.
- [99] D. Mozaffarian, E. J. Benjamin, A. S. Go, *et al.*, «Heart Disease and Stroke Statistics—2015 Update», *Circulation*, vol. 131, n.º 4, pp. e29-e322, ene. 2015, doi: 10.1161/CIR.000000000000152.
- [100] V. Kilintzis, I. Chouvarda, N. Beredimas, P. Natsiavas, y N. Maglaveras, «Supporting integrated care with a flexible data management framework built upon Linked Data, HL7 FHIR and ontologies», *J. Biomed. Inform.*, vol. 94, pp. 103179-103179, jun. 2019, doi: 10.1016/j.jbi.2019.103179.

- [101] J. J. Cimino y G. D. Fiol, «Infobuttons and Point of Care Access to Knowledge», en *Clinical Decision Support*, Elsevier Inc., 2007, pp. 345-371. doi: 10.1016/B978-012369377-8/50017-9.
- [102] M. Alfano, B. Lenzitti, G. Lo Bosco, C. Muriana, T. Piazza, y G. Vizzini, «Design, development and validation of a system for automatic help to medical text understanding», *Int. J. Med. Inf.*, vol. 138, jun. 2020, doi: 10.1016/j.ijmedinf.2020.104109.
- [103] A. Chetta, J. M. Carrington, y A. G. Forbes, «Augmenting EHR interfaces for enhanced nurse communication and decision making», en *ACM International Conference Proceeding Series*, New York, New York, USA, oct. 2015, vol. 25-October, pp. 1-6. doi: 10.1145/2836034.2836038.
- [104] S. Boytcheva, G. Angelova, Z. Angelov, D. Tcharaktchiev, y V. Vodenicharov, «Enrichment of EHR with linked open data for risk factors identification», en *ACM International Conference Proceeding Series*, New York, New York, USA, jun. 2019, pp. 84-90. doi: 10.1145/3345252.3345290.
- [105] N. G. Weiskopf, A. M. Cohen, J. Hannan, T. Jarmon, y D. A. Dorr, «Towards augmenting structured EHR data: a comparison of manual chart review and patient self-report», *AMIA Annu. Symp. Proc. AMIA Symp.*, vol. 2019, pp. 903-912, 2019.
- [106] «The Linking Open Data cloud diagram». [En línea]. Disponible en: <http://lod-cloud.net/>
- [107] C. Bizer, T. Heath, K. Idehen, y T. Berners-Lee, «Linked data on the web (LDOW2008)», en *Proceeding of the 17th International Conference on World Wide Web 2008, WWW'08*, New York, New York, USA, 2008, pp. 1265-1266. doi: 10.1145/1367497.1367760.
- [108] D. S. Wishart, C. Knox, A. C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, y J. Woolsey, «DrugBank: a comprehensive resource for in silico drug discovery and exploration», *Nucleic Acids Res.*, vol. 34, n.º 90001, pp. D668-D672, ene. 2006, doi: 10.1093/nar/gkj067.
- [109] U. Shah, T. Finin, A. Joshi, R. S. Cost, y J. Mayfield, «Information retrieval on the semantic web», en *International Conference on Information and Knowledge Management, Proceedings*, 2002, pp. 461-468. doi: 10.1145/584792.584868.
- [110] «SPARQL Query Results XML Format (Second Edition)». [En línea]. Disponible en: <https://www.w3.org/TR/rdf-sparql-XMLres/>
- [111] UPV. Biomedical Informatics Group, «LinKEHR Platform». [En línea]. Disponible en: <http://www.linkehr.com>
- [112] «Describing Linked Datasets with the VOID Vocabulary». [En línea]. Disponible en: <https://www.w3.org/TR/void/>
- [113] P. L. Whetzel, N. F. Noy, N. H. Shah, P. R. Alexander, C. Nyulas, T. Tudorache, y M.

- A. Musen, «BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications.», *Nucleic Acids Res.*, vol. 39, n.º Web Server issue, pp. W541-5, jul. 2011, doi: 10.1093/nar/gkr469.
- [114] «NCBO BioPortal - Repository of biomedical ontologies». [En línea]. Disponible en: <http://bioportal.bioontology.org/>
- [115] «SKOS Simple Knowledge Organization System Reference». [En línea]. Disponible en: <https://www.w3.org/TR/skos-reference/>
- [116] «SKOS Simple Knowledge Organization System Namespace Document - HTML Variant, 18 August 2009 Recommendation Edition». <https://www.w3.org/2009/08/skos-reference/skos.html#exactMatch> (accedido 10 de agosto de 2021).
- [117] «Bio2RDF - Linked Data for the Life Sciences». [En línea]. Disponible en: <http://bio2rdf.org/>
- [118] Spanish National Health System, «Resources for clinical modeling (Archetypes)». [En línea]. Disponible en: https://www.msssi.gob.es/en/profesionales/hcdsns/areaRecursosSem/Rec_mod_clinico_arquetipos.htm
- [119] J. Miñarro-Gimenez, M. Madrid, J. Fernandez-Breis, *et al.*, «OGO: an ontological approach for integrating knowledge about orthology», *BMC Bioinformatics*, vol. 10, n.º Suppl 10, pp. S13-S13, 2009, doi: 10.1186/1471-2105-10-S10-S13.
- [120] J. A. Maldonado, M. Marcos, J. T. Fernández-Breis, V. M. Giménez-Solano, M. del C. Legaz-García, y B. Martínez-Salvador, «CLIN-IK-LINKS: A platform for the design and execution of clinical data transformation and reasoning workflows», *Comput. Methods Programs Biomed.*, vol. 197, pp. 105616-105616, dic. 2020, doi: 10.1016/j.cmpb.2020.105616.
- [121] S. T. Konstantinidis, P. E. Kummervold, L. F. Luque, y L. K. Vognild, «A proposed framework to enrich norwegian EHR system with health-trusted information for patients and professionals», en *Studies in Health Technology and Informatics*, 2015, vol. 213, pp. 149-152. doi: 10.3233/978-1-61499-538-8-149.
- [122] F. Khaliq y S. A. Khan, «An FHIR-based framework for consolidation of augmented EHR from hospitals for public health analysis», abr. 2019. doi: 10.1109/ICAICT.2017.8687289.
- [123] M. Cheatham, A. Nikolov, Z. Zhang, A. Lisa Gentile, E. Blomqvist, I. Augenstein, y F. Ciravegna, «Semantic Web 0 (2015) 0 0 IOS Press An Unsupervised Data-driven Method to Discover Equivalent Relations in Large Linked Datasets», [En línea]. Disponible en: <http://www.semantic-web-journal.net/system/files/swj1052.pdf>
- [124] S. Issa, F. Hamdi, y S. S. said Cherfi, «Enhancing the Conciseness of Linked Data by

- Discovering Synonym Predicates», en *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, ago. 2019, vol. 11775 LNAI, pp. 739-750. doi: 10.1007/978-3-030-29551-6_65.
- [125] «Good practice for radiological reporting. Guidelines from the European Society of Radiology (ESR)», *Insights Imaging*, vol. 2, n.º 2, pp. 93-96, feb. 2011, doi: 10.1007/s13244-011-0066-7.
- [126] L. Martí-Bonmatí, «Radiomics and imaging biomarkers in observational clinical studies with retrospective data», *An. RANM*, vol. 136, n.º 01, pp. 34-42, may 2019, doi: 10.32440/ar.2019.136.01.rev07.
- [127] I. Radiology Technical Committee, «Technical Framework Supplement Management of Radiology Report Templates». [En línea]. Disponible en: <http://ihe.net/Profiles>.
- [128] J. Dickey, *Write Modern Web Apps with the MEAN Stack: Mongo, Express, AngularJS, and Node.js*. 2015.
- [129] «T-Rex». <https://t-rex.rsna.org/worklist.jsp> (accedido 8 de abril de 2022).
- [130] D. Pinto dos Santos, S. Brodehl, B. Baeßler, G. Arnhold, T. Dratsch, S. H. Chon, P. Mildenerger, y F. Jungmann, «Structured report data can be used to develop deep learning algorithms: a proof of concept in ankle radiographs», *Insights Imaging*, vol. 10, n.º 1, pp. 1-8, dic. 2019, doi: 10.1186/s13244-019-0777-8.
- [131] L. Martí-Bonmatí, Á. Alberich-Bayarri, y A. Torregrosa, «El informe radiológico. Estructura, estilo y contenido», *Radiología*, p. S0033833822000510, mar. 2022, doi: 10.1016/j.rx.2022.01.013.
- [132] R.- ASALE y RAE, «aumentar | Diccionario de la lengua española», «*Diccionario de la lengua española*» - Edición del Tricentenario. <https://dle.rae.es/aumentar> (accedido 6 de abril de 2022).
- [133] R.- ASALE y RAE, «enriquecer | Diccionario de la lengua española», «*Diccionario de la lengua española*» - Edición del Tricentenario. <https://dle.rae.es/enriquecer> (accedido 6 de abril de 2022).
- [134] Y. Wang, Z. Yao, y L. Liu, «Integration of DICOM and openEHR standards», Lake Buena Vista, Florida, mar. 2011, p. 79670T. doi: 10.1117/12.878360.
- [135] P. Moreno y G. Bastidas, «Propuesta metodológica de intercambio electrónico de información clínica basada en estándares de telemedicina/Methodological proposal of electronic data interchange and clinical information based on telemedicine standards», *KnE Eng.*, pp. 206-231, ene. 2020, doi: 10.18502/keg.v5i2.6237.
- [136] K. P. Lee y J. Hu, «XML Schema Representation of DICOM Structured Reporting.», *J. Am. Med. Inform. Assoc. JAMIA*, vol. 10, n.º 2, pp. 213-23, 2003, doi: 10.1197/jamia.m1042.
- [137] R. Isele Umbrich, J. ., Bizer, C. ., & Harth, A., «LDspider: An open-source crawling

framework for the Web of Linked Data.», en *In Proceedings of the 2010 International Semantic Web Conference (ISWC 2010)*, 2010, pp. 29-32. Vol. 658.

[138] «Gruff – New Browser Based Version | AllegroGraph». [En línea]. Disponible en: <https://allegrograph.com/products/gruff/>

Anexo I. Plantillas de informe radiológico estructurado

Algunas de las plantillas de IRE más relevantes desarrolladas en esta tesis se muestran a continuación. Las plantillas están organizadas en cuatro secciones: “Información Clínica”, “Información Técnica”, “Hallazgos” y “Conclusión”. La sección de datos demográficos del paciente se ha omitido en todas las plantillas para mayor claridad de las imágenes. La primera sección sirve para describir antecedentes clínicos del paciente con relevancia para el caso informado. A continuación, en “Información Técnica” se recogen los detalles sobre el procedimiento y el tipo de adquisición (por ejemplo, las secuencias adquiridas o el tipo de contraste utilizado). La sección de “Hallazgos” registra la información del análisis radiológico. Incluye campos para el diagnóstico cualitativo y, opcionalmente, resultados de SADC. Por último, en la sección “Conclusión” el radiólogo puede incluir observaciones adicionales o su diagnóstico final del caso. Estas plantillas cuentan con la aprobación del panel consultor de librerías de plantillas TLAP de la RSNA, y han sido desarrolladas en colaboración con un comité de expertos formado por radiólogos e ingenieros del Hospital Universitari i Politènic La Fe de Valencia.

Resonancia magnética de cerebro

Informe radiológico - RM Cerebro

Información Clínica

Información Técnica

RM Cerebro	<input type="text"/>	Técnica	<input type="text"/>
Comparación	<input type="text" value="Ninguna."/>	Contraste IV	<input type="text"/>

Hallazgos

Espacio subaracnoideo	<input type="text" value="Normal en tamaño y morfología para la edad"/>	Hemorragia intracraneal	<input type="text" value="Ninguna."/>
Sistema ventricular	<input type="text" value="Normal en tamaño y morfología para la edad"/>	Cisternas basales	<input type="text" value="Normal."/>
Parénquima cerebral	<input type="text" value="Normal."/>	Cambio de línea media	<input type="text" value="Normal."/>
Cerebelo	<input type="text" value="Normal."/>	Tronco encefálico	<input type="text" value="Normal."/>
Bóveda craneal	<input type="text" value="Normal."/>	Sistema vascular	<input type="text" value="Vacíos de flujo normal, sin presencia de"/>
Senos paranasales y celdillas mastoideas	<input type="text" value="Limpios."/>	Órbita visible	<input type="text" value="Normal."/>
Columna cervical visible	<input type="text" value="Normal."/>	Silla turca	<input type="text" value="Normal."/>
Base del cráneo	<input type="text" value="Normal."/>	Médula	<input type="text" value="Normal."/>

Conclusión

Figura 30. Formulario de IRE para IRM de cerebro. Se ha implementado a partir de la plantilla para IRM de cerebro de la iniciativa Radreport, aprobada por el TLAP.

Resonancia magnética de columna completa sin contraste

Informe radiológico - RM de la columna completa sin contraste

Información Clínica

Información técnica

Procedimiento

Las imágenes de RM de la columna cervical, torácica y lumbar se

Comparación

Ninguna.

Hallazgos

Anatomía

7 cuerpos vertebrales cervicales, 12 tor:

Médula espinal

Normal.

El cono termina en el nivel

T12-L1.

Discos

Normal.

Huesos

La altura del cuerpo vertebral y la alineación

Señal medular

Normal.

Tejido blando

Normal.

Columna cervical

C2-C3

No hay canal ni estenosis foraminal.

C3-C4

No hay canal ni estenosis foraminal.

C4-C5

No hay canal ni estenosis foraminal.

C5-C6

No hay canal ni estenosis foraminal.

C6-C7

No hay canal ni estenosis foraminal.

C7-T1

No hay canal ni estenosis foraminal.

Columna torácica

No hay canal ni estenosis foraminal.

Columna lumbar

T12-L1

No hay canal ni estenosis foraminal.

L1-L2

No hay canal ni estenosis foraminal.

L2-L3

No hay canal ni estenosis foraminal.

L3-L4

No hay canal ni estenosis foraminal.

L4-L5

No hay canal ni estenosis foraminal.

L5-S1

No hay canal ni estenosis foraminal.

Otros hallazgos

Ninguno.

Conclusión

RM normal de la columna vertebral completa.

Figura 31. Formulario de IRE para IRM de columna completa sin contraste. Se ha implementado a partir de la plantilla para IRM de columna completa sin contraste de la iniciativa Radreport, aprobada por el TLAP.

Resonancia magnética de rodilla

Informe radiológico - RM Rodilla

Información Clínica

Dolor en rodilla.

Información Técnica

RM Rodilla

Se obtuvieron imágenes de RM transversal, sagital y coronal de la rodilla.

Selecciona región

Derecha

Comparación

Ninguna.

Hallazgos

MENISCOS

Menisco medial

Normal.

Menisco lateral

Normal.

LIGAMENTOS

Ligamentos cruzados

Normal.

Ligamento colateral medial

Componentes superficiales y profundos

Ligamento colateral lateral

Normal.

Estructuras posterolaterales

Normal.

MECANISMO EXTENSOR

La porción distal de los cuádriceps y el tendón rotuliano están intactos. La rótula se encuentra dentro del surco femoral. No hay disrupción retinacular.

DERRAME / BURSAS / QUISTES

No hay derrame articular. No hay quiste de Baker.

HUESO Y CARTÍLAGO

Huesos

No se observa ninguna fractura, reacción

Compartimento femoropatelar

No hay enfermedad del cartílago.

Compartimento medial

No hay enfermedad del cartílago.

Compartimento lateral

No hay enfermedad del cartílago.

Conclusión

1.

Menisco normal.

2.

No hay desgarros en el ligamento cruzado o colateral.

3.

No hay anomalías del hueso o cartílago.

Figura 32. Formulario de IRE para IRM de rodilla. Se ha implementado a partir de la plantilla para IRM de rodilla de la iniciativa Radreport, aprobada por el TLAP.

Resonancia magnética de cadera

Informe radiológico - RM Cadera

Información Clínica

Información técnica

Procedimiento	Se obtuvieron imágenes de densidad de protones transversal,	Lateralidad	Cadera derecha
Comparación	Ninguna.		

Hallazgos

Alineamiento	Normal. Sin subluxación		
Anatomía del pinzamiento femoroacetabular			
Pinzamiento de retroversión	Ausente	Sobrecobertura	Ausente
Os acetabuli	Ausente		
Cam			
Ángulo cabeza-cuello	Normal.	Lesión fibroquística del cuello femoral	Ninguna.
Displasia	Ninguna.	Derrame articular	Ninguno.
Labrum	Normal. Sin desgarro del labrum.		
Cartílago			
Cartilago femoral	Normal.	Cartilago acetabular	Normal.
Capsula y ligamentos	Normal.		
Músculos, tendones y entesis			
Flexores	Normal.	Extensores	Normal.
Abductores	Normal.	Aductores	Normal.
Rotadores	Normal.	Isquiotibiales	Normal.
Otros hallazgos			
Vasos	Normal.	Nervios	Normal.
Tejido blando	Normal.	Viscera	Normal. No se aprecian anomalías en la

Conclusión

Figura 33. Formulario de IRE para IRM de cadera. Se ha implementado a partir de la plantilla para IRM de cadera de la iniciativa Radreport, aprobada por el TLAP.

Tomografía computarizada de abdomen y pelvis

Informe radiológico - TC abdomen y pelvis

Información Clínica

Información técnica

Estudio del TC de abdomen y pelvis

con contraste IV. ▾

Indicaciones

Técnica

Rutina - Se obtuvieron imágenes de exploración del corazón a través del abdomen y la pelvis después de la administración del protocolo de ma ▾

Comparación

No hay TCs abdominales o pélvicos previos en el sistema.

Hallazgos

Bases pulmonares

Hígado

No hay lesiones de masa focal. No hay dilatación ductal biliar intrahepática.

Vesícula biliar

Normal - La apariencia del TC es cc ▾

Bazo

Corriente.

Pancreas

Apariencia normal del TC

Adrenales

Sin nódulos.

Riñones

Sin lesiones sólidas, cálculos, ni hidroureteronefrosis.

Mesentero / Peritoneo

Corriente.

Intestinos

Normal-Apéndice normal identificado ▾

Comentarios

Sin dilatar.

Pelvis

Hombre joven - La próstata y las ves ▾

Comentarios

Ventanas óseas

Sin anomalías óseas agresivas.

Vasculatura

Normal - Normal, sin aneurisma o er ▾

Tejidos blandos

Normal - Corriente. ▾

Otros hallazgos

Conclusiones

No hay enfermedad maligna o metastásica en el abdomen o la pelvis.

Figura 34. Formulario de IRE para TAC de abdomen y pelvis. Se ha implementado a partir de la plantilla para TAC de abdomen y pelvis de la iniciativa Radreport, aprobada por el TLAP.

Tomografía computarizada de cuello

Informe radiológico - TC Cuello

Información Clínica

Información técnica

Procedimiento

Comparación

Ninguna.

Hallazgos

Órbitas, senos
paranasales y bases
del cráneo

Normal.

Nasofaringe

Normal.

Cuello suprahioides

Orofaringe, cavidad oral, espacio parafaríngeo

Cuello infrahioides

Laringe, hipofaringe y supraglotis normales

Tiroides

Normal.

Entrada torácica

Ápices pulmonares y plexo braquial normales

Ganglios linfáticos

Normal. No existe linfadenopatía.

Estructuras
vasculares

Normal.

Otros hallazgos

Ninguno.

Conclusión

Estudio normal.

Figura 35. Formulario de IRE para TAC de cuello. Se ha implementado a partir de la plantilla para TAC de cuello de la iniciativa Radreport, aprobada por el TLAP.

Tomografía computarizada de huesos faciales

Informe radiológico - TC huesos faciales

Información Clínica	
<input type="text"/>	

Información técnica	
Procedimiento	<input type="text"/>
Comparación	<input type="text" value="Ninguna."/>

Hallazgos	
Huesos faciales	<input type="text" value="No hay fractura ósea facial evidente."/>
Otras estructuras	<input type="text" value="Los senos paranasales visibles están bien aireados. No hay"/>
Otros hallazgos	
<input type="text"/>	

Conclusión
<input type="text" value="No hay evidencia de fracturas o dislocaciones en los huesos faciales."/>

Figura 36. Formulario de IRE para TAC de huesos faciales. Se ha implementado a partir de la plantilla para TAC de huesos faciales de la iniciativa Radreport, aprobada por el TLAP.

Tomografía computarizada de senos paranasales

Informe radiológico - Senos paranasales

Información Clínica

Información técnica

Procedimiento

Comparación

Hallazgos

Senos paranasales limpios. No hay evidencia del nivel de flujo de aire.

Otros hallazgos

Conclusión

Estudio negativo. No hay evidencia de sinusitis.

Figura 37. Formulario de IRE para TAC de senos paranasales. Se ha implementado a partir de la plantilla para TAC de senos paranasales de la iniciativa Radreport, aprobada por el TLAP.

Tomografía computarizada de embolia pulmonar de tórax

Informe radiológico - TC Embolia pulmonar de tórax

Información Clínica

Información Técnica

Contraste intravenoso (agente y volumen)

Comparación

Ninguna.

Hallazgos

Calidad diagnóstica

Adecuada

Embolia pulmonar

Ninguna.

Dilatación del ventrículo derecho

Ninguna.

Tronco de la arteria pulmonar

De calibre normal.

Parénquima pulmonar

Normal.

Derrame pleural

Ninguno.

Vías respiratorias centrales

Normal.

Adenopatías

Ninguna.

Corazón y grandes vasos

Normal.

Abdomen superior

Normal.

Huesos

Normal.

Otros

Conclusión

No hay embolismo pulmonar agudo o crónico.

Figura 38. Formulario de IRE para TAC de embolia pulmonar de tórax. Se ha implementado a partir de la plantilla para TAC de embolia pulmonar de tórax de la iniciativa Radreport, aprobada por el TLAP.

Radiografía de rodilla

Informe radiológico - Radiografía de rodilla

Información Clínica

Información técnica

Lateralidad	<input type="text" value="Rodilla izquierda."/>	Número de proyecciones	<input type="text"/>
Procedimiento	<input type="text"/>	Comparación	<input type="text" value="Ninguna."/>

Hallazgos

Huesos	<input type="text" value="Sin presencia de fracturas o dislocacion"/>	Articulaciones	<input type="text" value="Los espacios articulares son normales."/>
Tejido blando	<input type="text" value="Normal."/>		

Otros hallazgos

Conclusión

Figura 39. Formulario de IRE para radiografía de rodilla. Se ha implementado a partir de la plantilla para radiografía de rodilla de la iniciativa Radreport, aprobada por el TLAP.

Radiografía de tórax con dos proyecciones

Informe radiológico - Radiografía de tórax - 2 proyecciones

Información Clínica

Información técnica

Procedimiento

Proyecciones postero-anterior y lateral del tórax.

Comparación

Ninguna.

Hallazgos

Pulmones

Los pulmones no presentan alteraciones.

Corazón

El corazón es de tamaño normal.

Mediastino

El mediastino está dentro de los límites normales.

Otros hallazgos

Conclusiones

Enfermedad no activa.

Figura 40. Formulario de IRE para radiografía de tórax con dos proyecciones. Se ha implementado a partir de la plantilla para radiografía de tórax con dos proyecciones de la iniciativa Radreport, aprobada por el TLAP.

Anexo II. Instituciones sanitarias y/o de I+D+i utilizando el sistema de informe radiológico estructurado

El sistema de informe radiológico estructurado descrito en esta tesis, se ha probado y puesto en funcionamiento con éxito en las siguientes instituciones sanitarias y/o de I+D+i:

1. "Equipo Juana Crespo". Clínica de fertilidad en Valencia, España.
2. "Hospital Clínic de Barcelona". Hospital universitario en Barcelona, España.
3. "Fundación Clínic de Barcelona". Fundación de investigación en Barcelona, España.
4. "Fundación Biomédica Galicia Sur". Fundación de investigación biomédica en Galicia, España.
5. "Instituto de Resonancia Magnética Dra. Guirado". Instituto de imagen de resonancia magnética en Barcelona, España.
6. "Instituto de Investigación Sanitaria La Fe". Instituto de investigación en Valencia, España.
7. "Institut Català de la Salut (ICS)". Instituto de salud en Barcelona, Spain.
8. "Fundación de Investigación HM Hospitales". Fundación de investigación en Madrid, España.
9. "MRIcons", Centro de imagen de resonancia magnética en Nicosia, Chipre
10. "Champalimaud Foundation". Fundación de investigación biomédica en Lisboa, Portugal.
11. "SALUS, Medicina y Gestión Sanitaria". Centros de diagnóstico por imagen en Murcia, España.
12. "Hospital Universitario de Canarias". Hospital universitario en Canarias, España.
13. "RadiewCare". Centro de diagnóstico por la imagen en Madrid, España.
14. "Hospital Germans Trias i Pujol". Hospital en Barcelona, España.
15. "Hospital Ramón y Cajal". Hospital en Madrid, España.
16. "Instituto Ramón y Cajal de Investigación Sanitaria". Instituto de investigación sanitaria en Madrid, España.
17. "Clínica DKF". Clínica de salud en Madrid, España.
18. "Fundación Del Hospital Provincial De Castellón". Fundación de investigación sanitaria en Castellón, España.
19. "Hospital Sant Joan de Deu". Hospital en Barcelona, España.
20. "Kurume University". Hospital universitario en Kurume, Japón.
21. "Universitat de Girona". Universidad en Girona, España.
22. "Laboratorio Clínico Diagnomed, S.A.". Laboratorio clínico en Concepción, Chile.
23. "Molecular Oncology Institute FIRC – IFOM". Instituto de investigación en Milán, Italia.

24. "Centro Uruguayo de Imagenología Molecular". Centro de imagen médica en Montevideo, Uruguay.
25. "Centro de Medicina Comparativa y Bioimagen de Cataluña". Centro de investigación en Barcelona, España.
26. "Fundació Privada per a la Recerca Sant Joan de Dèu". Fundación de investigación en Barcelona, España.
27. "Wanfang Hospital". Hospital en Taipéi, Taiwán.
28. "Mackay Memorial Hospital". Hospital en Taipéi, Taiwán.
29. "Azienda USL di Reggio Emilia, IRCCS". Hospital en Reggio Emilia, Italia.
30. "Hospital Universitari i Politècnic La Fe". Hospital Universitario y Politécnico en Valencia, España.
31. "Kaohsiung Veteran General Hospital". Hospital en Kaohsiung, Taiwán.
32. "IEC Core Lab". Laboratorio central de imagen médica para ensayos clínicos en Barcelona, España.
33. "Imaging Endpoints". Laboratorio central de imagen médica para ensayos clínicos en Massachusetts, Estados Unidos.
34. "Ferring Pharmaceuticals". Compañía farmacéutica en Copenhague, Dinamarca.
35. "Complejo Hospitalario Universitario de Vigo". Hospital universitario en Vigo, España.
36. "Global Institute for Research". Instituto de investigación para ensayos clínicos en Arizona, Estados Unidos.
37. "Highlight Therapeutics". Compañía de investigación para ensayos clínicos en Valencia, España.

