

# Análisis de Fuentes de Datos Públicas Relevantes para la Historia Clínica Electrónica (HCE)

## *Analysis of the Relevant Public Data Sources for Electronic Health Record (EHR)*

Isabel de la Torre Díez<sup>1</sup>

<sup>1</sup>Profesora de la Universidad de Valladolid (España).

### Resumen / Abstract

**Resumen.** *Las fuentes de datos públicas son muy útiles para el acceso a información a través de la Historia Clínica Electrónica (HCE). Todas ellas están disponibles vía Web de forma gratuita. En este artículo se han analizado 43 de ellas. También, se ha incluido la información más relevante de las mismas y el acceso Web para acceder a su repositorio digital. Se ha podido observar que la mayoría de ellas son de origen estadounidense y canadiense.*

**Abstract.** *Public data sources are useful for accessing to information through Electronic Health Record (EHR). All of them are available via the Web and free. In this article, we have examined 43 of them. Their most relevant information and their Web have been included. It has been observed that most of them are American and Canadian.*

## 1. Introducción

Los repositorios de datos también conocidos como Archivos de Acceso Abierto (Open Access) son archivos digitales accesibles a través de Internet. Reúnen la producción intelectual de una institución o de una disciplina. Desde hace tiempo los repositorios de datos digitales se han convertido en un tema importante. En el campo de la medicina también existen gran cantidad de repositorios de datos clínicos y biomédicos.

Las fuentes de datos nacen para cubrir las necesidades informativas de determinadas ciencias. En cualquier investigación es necesaria información científica. En el campo de la biomedicina los usuarios necesitan contar rápidamente con información novedosa, muy precisa, actualizada, de calidad, y a ser posible filtrada. Las fuentes de información en Biomedicina son muy amplias y diversas, abarcan desde las fuentes primarias (por ejemplo, revistas y monografías), las fuentes secundarias (bibliografías, catálogos, bases de datos, etc.) y de referencia.

En este informe se analizan fuentes de datos públicas existentes (bases de datos) y accesibles vía Web, muy útiles en la implantación de sistemas de Historia Clínica Electrónica (HCE), partiendo de información extraída de [1].

## 2. Fuentes de Datos Públicas en Medicina

A continuación, pasamos a describir repositorios de datos públicos existentes en la actualidad en el campo de la medicina.

### 2.1. ADME databases

Estas bases de datos se recogen cuidadosamente de la literatura. En un futuro cercano, más bases de datos

se agregarán. Se trata de bases de datos de interés para quienes trabajan en el descubrimiento de medicamentos para la evaluación comparativa de los resultados de experimentos, validar la exactitud de los modelos predictivos ADME existentes y construir nuevos modelos predictivos utilizando un conjunto de datos de entrenamiento.

Incluyen siete bases de datos:

1. Sobre la solubilidad del agua.
2. Sobre permeabilidad Caco-2.
3. Sobre permeabilidad hematoencefálica.
4. Sobre inhibidor P-gp.
5. Sobre absorción oral.
6. De bio-disponibilidad oral.
7. Bases de datos combinadas.

El responsable del desarrollo de modelos predictivos ADME, bases de datos y librería de Objetos y plantillas relevantes es el Dr. Tingjun Hou, del Institute of Functional Nano & Soft Materials (FUNSOM), de la Universidad de Soochow en China [2].

### 2.2. BCCA Cancer Drug Manual

El principal objetivo del Manual de drogas del cáncer © (CDM) es proporcionar una fuente de referencia oportuna, precisa y objetiva para profesionales sanitarios y cuidadores de pacientes con cáncer. El manual contiene información concisa y evaluativa sobre medicamentos para los tratamientos de cáncer. La información se presenta utilizando un formato normalizado por temas de coherencia y legibilidad. Los documentos están escritos por farmacéuticos de Oncología, a menudo con el apoyo de médicos expertos. Estos documentos son luego revisados y aprobados por un Comité editorial multidisciplinario de farmacéuticos de Oncología, oncólogos médicos y enfermeras de Oncología, para garantizar la calidad, integridad y relevancia clí-

nica de la información. El Manual de drogas del cáncer<sup>®</sup> es un recurso de información general de las drogas. Para cuestiones relacionadas con terapias, el lector tiene que dirigirse a los protocolos de quimioterapia de la Agencia de cáncer de BC, a las directrices de gestión de cáncer y a la literatura más actual. Cualquier profesional que utilice información del Manual de drogas del cáncer para proporcionar tratamiento para los pacientes será el responsable de verificar las dosis, proporcionar las prescripciones y administrar los medicamentos descritos en el manual según normas aceptables existentes [3].

### 2.3. CAS

La base de datos CAS (Chemical Abstracts Service) ofrece información sobre química y otras disciplinas científicas, incluyendo las ciencias biomédicas, química, ingeniería, ciencia de los materiales y ciencias agrícolas, entre otras. Las bases de datos de CAS contienen más de 67 millones únicas sustancias orgánicas e inorgánicas químicos, tales como aleaciones, compuestos de coordinación, minerales, mezclas de polímeros y sales y, más de 63 millones de secuencias [4]. CAS es una división de American Chemical Society (ACS), y produce Chemical Abstracts y productos relacionados. Está localizada en Columbus, Estados Unidos.

Proporciona la mayor base de datos revelada públicamente sobre Química, y la hace accesible buscando y creando software que provee enlaces a la literatura y las patentes originales. En el año 2007, el Servicio de Chemical Abstracts estaba diseñado por el ACS National Historical Chemical Landmark, en reconocimiento a su significado en la mayor parte de las investigaciones científicas dentro de la Química.

### 2.4. ChEBI

Se trata de una base de datos de entidades moleculares (cualquier producto natural o de síntesis usado para intervenir en el proceso de organismos vivos). Está centrada en pequeños componentes químicos. También es un diccionario gratuito de entidades moleculares centrado en compuestos químicos 'pequeños'. No están incluidas en esta base de datos moléculas directamente relacionadas con el genoma como son los péptidos, las proteínas y los ácidos nucleicos [5].

### 2.5. ChemPDB

La base de datos ChemPDB está formada por rayos x o estructuras RMN (espectroscopia mediante resonancia magnética nuclear de proteínas) 3D en formato PDB; conformaciones de ligamentos que están vinculadas con proteínas. Al tener una estructura tridimensional, la base de datos no necesita generación de conformación.

El servicio de diccionario de ChemPDB proporciona acceso Web al diccionario componente químico de la wwPDB cuando éste es cargado en la base de datos PDBe en EBI. Este diccionario es parte de la información de referencia del núcleo de la base de datos relacional de PDBe [6].

### 2.6. DailyMed

DailyMed proporciona información de calidad sobre los medicamentos comercializados. Esta información incluye etiquetas de FDA (prospectos). Este sitio Web proporciona a los proveedores de salud y al público en general un recurso de consulta y descarga estándar, integral y actualizada, de contenido de medicación y etiquetado como se encuentra en los prospectos de medicamentos [7]. La Biblioteca Nacional de Medicina (de las siglas en inglés NLM, National Library of Medi-

cine) proporciona un servicio público y no acepta publicidad. Dispone de versión móvil.

## 2.7. DBpedia

DBpedia es un proyecto para la extracción de datos de Wikipedia para proponer una versión de Web semántica. Está interconectada con GeoNames, CIA World Factbook, Proyecto Gutenberg, Eurostat entre otros. En esta base de datos se describen 3.640.000 entidades, entre ellas al menos 416.000 personas, 526.000 lugares, 106.000 álbumes de música y 60.000 películas y contiene 2,724,000 enlaces a imágenes, 6,300,000 enlaces a páginas externas, 6,200,000 enlaces a datasets externos y 740,000 categorías Wikipedia, entre ellas de medicina y bioingeniería [8]. La información se almacena con el RDF (Resource Description Framework), podemos hacer consultas a la base de datos a través de SPARQL.

## 2.8. dbSNP

La base de datos Single Nucleotide Polymorphism (dbSNP) es un archivo público y de uso gratuito sobre variación genética de diferentes especies desarrollado y alojado en el Centro Nacional para la Información de Biotecnología (el NCBI, National Center for Biotechnology Information) en colaboración con el Instituto de Investigación Nacional del Genoma Humano (el NHGRI, National Human Genome Research Institute). Aunque el nombre de la base de datos implica una colección de una sola clase de polimorfismos (por ejemplo, los polimorfismos nucleótidos simples), contiene también un rango de variación molecular como: borrado e inserción de polimorfismos, marcas microsatélites, polimorfismos multinucleidos, secuencias heterogéneas, otras variantes. Esta base de datos acepta polimorfismos neutros, polimorfismos correspondien-

tes a conocidos fenotipos, y regiones sin variaciones. Fue creado en Septiembre de 1998 to sustituir a GenBank, la colección de NCBI de ácidos nucleicos y secuencias de proteínas [9-11].

## 2.9. Diseasome

Se trata de una base de datos integrada sobre genes, variación genética y enfermedades. A través de un mapa interactivo, un póster o libro se representa el mapa de las enfermedades humanas y cómo están relacionadas entre sí.

Debido a que el polimorfismo de un solo nucleótido o SNP (Single Nucleotide Polymorphism) puede causar enfermedades, la relación entre la variación genética y las enfermedades se ha convertido en un tema clave. A pesar de esta necesidad, hay muy pocas bases de datos de variaciones genéticas asociadas a enfermedades. La base de datos Diseasome fue desarrollada para asociar información de variación genética con enfermedades.

Proporciona formas semiautomáticas de derivar la lista de candidatos SNPs a evaluarse en experimentos biológicos epidemiológicos o moleculares para estudios de asociación de la enfermedad. Actualmente, contiene 14.674 registros sobre variación genética y 109.715 sobre genes relacionados con enfermedades [12].

## 2.10. DrugBank

La base de datos DrugBank es un recurso bioinformático y químicoinformático que combina drogas detalladas (por ejemplo: químicas, farmacológicas y farmacéuticas) con dianas farmacológicas (por ejemplo: secuencias, estructuras) de información.

La base de datos contiene 6711 entradas de medicamentos incluyendo 1447 medicamentos de moléculas

pequeñas aprobadas por el FDA, 131 medicamentos de proteínas/péptidos aprobados por el FDA, 85 nutracéuticos y 5080 medicamentos experimentales. También, 4227 proteínas no redundantes como son secuencias de encimas, etc. son enlazadas a estas entradas de medicamentos. Cada entrada DrugCard contiene más de 150 campos de datos con la mitad de información relativa a información química o de medicamentos y la otra mitad relativa a datos de proteínas o dianas farmacológicas. DrugBank es mantenida por David Wishart, del Departamento de Ciencias de la Computación y Biológicas de la Universidad de Alberta en Canadá, y por el Centro de Innovación Metabólica<sup>1</sup> canadiense formado por comunidad científica y por empresas líderes en tecnología metabólica [13].

## 2.11. Drugs.com

Drugs.com es la más popular, amplia y actualizada fuente de información de medicamentos on-line en Estados Unidos. Proporciona datos de forma segura e independiente de más de 24000 medicamentos y productos naturales. Dispone de herramientas útiles, registros de medicamentos personales, aplicaciones móviles, etc. [14].

## 2.12. FDA

La FDA (Food and Drug Administration: Agencia de Alimentos y Medicamentos o Agencia de Drogas y Alimentos) es la agencia del gobierno de los Estados Unidos responsable de la regulación de alimentos

(tanto para personas como para animales), suplementos alimenticios, medicamentos (humanos y veterinarios), cosméticos, aparatos médicos (humanos y animales), productos biológicos y derivados sanguíneos [15].

## 2.13. GenBank

GenBank<sup>®</sup> es la base de datos de secuencias genéticas del NIH (Instituto Nacional de Salud de los Estados Unidos), una colección con notaciones de todas las secuencias de ADN disponibles de carácter público [16]. Existen aproximadamente 126,551,501,141 bases en 135,440,924 secuencias en las divisiones tradicionales de GenBank y 191,401,393,188 bases en 62,715,288 secuencias en la división WGS desde Abril de 2011.

Las notas de los releases completos para la version actual de GenBank están disponibles a través del sitio de NCBI vía FTP. Cada dos meses se hace un nuevo release. GenBank es parte de la Colaboración de la Bases de Datos de Secuencia Nucleótida Internacional, con el Banco de Datos de ADN de Japón (DDBJ) y el Laboratorio de Biología Molecular Europeo (EMBL). Estas tres organizaciones intercambian información de forma diaria [17].

## 2.14. GeneID

Gene integra información de un gran rango de especies. Un historial de la base de datos puede incluir nomenclatura, Secuencias de Referencia (RefSeqs), mapas, destinos, variaciones, fenotipos, y enlaces a genomas, fenotipos, y otras fuentes específicas [18].

<sup>1</sup>Grupo de ciencias y técnicas dedicadas al estudio completo del sistema constituido por el conjunto de moléculas que forman los intermediarios metabólicos, metabolitos, hormonas y otras moléculas señal, y también los metabolitos de carácter secundario, que se pueden encontrar en un sistema biológico. Grupo de ciencias y técnicas dedicadas al estudio completo del sistema constituido por el conjunto de moléculas que forman los intermediarios metabólicos, metabolitos, hormonas y otras moléculas señal, y también los metabolitos de carácter secundario, que se pueden encontrar en un sistema biológico.

## 2.15. Gene Ontology

El proyecto Gene Ontology es una de las mayores iniciativas bioinformáticas con el objetivo de estandarizar la representación de genes y los atributos de productos genéticos a través de especies y bases de datos. El proyecto suministra un vocabulario controlado de términos que describen las características de los productos de genes y los datos de anotación de esos productos genéticos [19].

## 2.16. HapMap

El proyecto internacional HapMap está formado por científicos y agencias de diferentes países como Canadá, China, Japón, Nigeria, Reino Unido y Estados Unidos de América. La idea es desarrollar una base de datos pública que ayude a los investigadores a encontrar genes asociados a enfermedades humanas [20]

## 2.17. Health Canada Drug Product Database (DPD)

DPD contiene información específica sobre medicamentos en uso en Canadá. La base de datos es gestionada por el Instituto de Salud de Canadá e incluye drogas biológicas y farmacéuticas humanas, productos desinfectantes y drogas veterinarias. Contiene aproximadamente 15,000 productos de diferentes compañías comerciales que han sido notificadas con el Instituto de Salud de Canadá [21].

## 2.18. HGNC

Son las siglas de HUGO Gene Nomenclature Committee. El Comité de Nomenclatura de Genes HUGO es la única autoridad que asigna nomenclatura estándar a genes humanos. HUGO es una organización dentro del

Proyecto de Genoma Humano cuyo principal objetivo es el mapeo de dicho genoma. Esta organización fue fundada en el año 1989 como internacional. Su Comité de Nomenclatura es el HGNC que asigna un único nombre y símbolo para cada uno de los genes humanos. Se trata de un comité cofundado por el Instituto Nacional de Investigación del Genoma Humano americano (US National Human Genome Research Institut-NHGRI) y el instituto inglés Wellcome Trust [22].

## 2.19. KEGG Compound

KEGG (Kyoto Encyclopedia of Genes and Genomes) (Enciclopedia de Genes y Genomas de Kyoto) Compound es una base de datos de estructura química para componentes metabólicos y otras sustancias químicas relevantes para sistemas biológicos. Es una colección de bases de datos en línea de genomas, rutas enzimáticas, y químicos biológicos. Mantiene 5 bases de datos principales: KEGG Atlas, KEGG Pathway, KEGG Genes, KEGG Ligand y KEGG BRITE.

La base de datos PATHWAY registra las redes de interacciones moleculares dentro de las células, y variantes de éstas específicas a organismos particulares. Cada entrada es identificada por el número C, y contiene varios enlaces a otras bases de datos KEGG y también a otras bases fuera de ésta. A partir de julio de 2011, KEGG ha cambiado a un modelo de suscripción y el acceso a través de FTP ya no es gratis [23].

## 2.20. Ligand Depot

Ligand Expo (denominado formalmente Ligand Depot) proporciona información estructural y química sobre pequeñas moléculas en las entradas de la PDB (Protein Data Bank). Las herramientas son suministradas para buscar en el diccionario PDB por componentes químicos, para identificar las entradas que contienen pe-

queñas partículas, y para descargar estructuras en 3D de los componentes de las pequeñas moléculas. También hay una herramienta para construir nuevas definiciones químicas desde los componentes químicos reportados en la PDB [24].

### 2.21. LinkedCT

El proyecto LinkedCT tiene por objeto publicar la primera fuente de datos Web semántica de acceso libre relative a información clínica. Su base de datos es generada (1) transformando las Fuentes de datos clínicos existentes al formato RDF, y (2) descubriendo los enlaces semánticos entre los historiales y otras fuentes de datos clínicos [25].

### 2.22. MediCare

La base de datos Medicare Coverage (MCD) contiene todas las Determinaciones de Cobertura Nacional denominadas: National Coverage Determinations (NCDs) y las locales: Local Coverage Determinations (LCDs), artículos locales y decisiones NCD propuestas. Se trata de un programa de cobertura de seguridad social administrado por el gobierno de Estados Unidos, que suministra atención médica a personas mayores de 65 años.

El programa también financia los programas de formación de médicos residentes en Estados Unidos. Medicare opera como un seguro de personas.

También incluye diversos tipos de políticas de cobertura nacional relativas a documentos, incluyendo los Análisis de Cobertura Nacional llamados: National Coverage Analyses (NCAs), Análisis de códigos para laboratorios: Coding Analyses for Labs (CALs), Desarrollo de la Evidencia de Cuidados médicos, documentos de guías de cuidados médicos, proceedings del Comité de

Consejos de Cobertura: Coverage Advisory Committee (MEDCAC) [26].

### 2.23. NIST WebBook

Base de Datos de Referencia Estándar del NIST Número 69. Este sitio electrónico proporciona datos termoquímicos, termofísicos y de energética de iones compilados por el NIST en el ámbito del Programa de Datos de Referencia Estándar (Standard Reference Data Program).

El National Institute of Standards and Technology (NIST) se esfuerza por proporcionar una copia de alta calidad de la base de datos y por verificar que los datos contenidos en esta sean seleccionados con base en elevados criterios científicos. Sin embargo, el NIST no da garantías a este efecto, y el NIST no se responsabiliza por cualquier daño que pueda resultar de errores u omisiones en la base de datos [27].

### 2.24. OMIM – Online Mendelian Inheritance in Man<sup>®</sup>

OMIM es una base de datos formada por un compendio de genes humanos y fenotipos genéticos. Cataloga todas las enfermedades conocidas con un componente genético, y cuando es posible, la asociación a los genes en el genoma humano. La base de datos está disponible en formato de libro denominado Inheritance in Man (MAM). La versión online se llama Online Mendelian Inheritance in Man' (OMIM), pudiendo ser accedida a través del motor de la base de datos de la Librería Nacional de Medicina llamada Entrez. El autor y editor es el doctor [Loui di termine y sus compañeros de la universidad Johns Hopkins, desarrollado para la WWW por la NCBI (National Center for Biotechnology Information). Está disponible en Internet desde 1987 [28].

## 2.25. Open Babel

Open Babel es una herramienta diseñada en varios idiomas que trata de elementos químicos. Es de software libre y permite buscar, convertir, analizar o almacenar datos desde modelado molecular, química, materiales de estado sólido, bioquímica u otras áreas [29].

## 2.26. PDB

Protein Data Bank [30]. Se trata de una base de datos de la estructura tridimensional de las proteínas y ácidos nucleicos. Están bajo el dominio público y se pueden usar libremente [31].

## 2.27. PDRhealth

PDRhealth™ es un portal Web de la red PDR ([www.pdrnetwork.net](http://www.pdrnetwork.net)), que incluye la Referencia la Es-critorio del Médico denominada: Physicians' Desk Reference (PDR) y la red de Anotación para el Cuidado de la Salud: Health Care Notification Network (HCNN). Este sitio ofrece explicaciones de forma agradable al usuario sobre los estados de enfermedades y las condiciones, y también sobre el uso efectivo de la prescripción y de la no prescripción de los medicamentos y de medicinas herbales.

Este sitio proporciona información clínica a los médicos sobre todo en Estados Unidos desde hace 60 años. El HCNN es el único servicio que suministra información de tipo electrónico a los médicos/as y enferme-ros/as sobre productos relativos a la seguridad del paciente y órdenes de la FDA. Dirigido por la organización sin ánimo de Lucro iHealth Alliance, el HCNN proporciona actualización sobre medicamentos y alertas de seguridad en el campo farmacéutico en los Es-

tados Unidos [32].

## 2.28. Pfam

Pfam es una amplia colección de alineamientos múltiples de secuencias y modelos ocultos de Márkov cubriendo buena parte de dominios de proteínas. Para cada familia en Pfam se pueden revisar las arquitecturas y organización de los dominios proteicos, examinar la distribución de especies, seguir enlaces a otras bases de datos, ver los alineamientos múltiples y las estructuras proteicas conocidas [33].

## 2.29. PharmGKB

PharmGKB es un recurso de conocimiento farmacogénico acompañado de información clínica y que incluye guías de dosis, niveles de medicamentos, asociaciones entre medicamentos y genes activos clínicamente, relaciones entre genotipos y fenotipos. Esta base de datos almacena y difunde información sobre el impacto de la variación genética humana en las respuestas de los medicamentos a través de algunas de las siguientes actividades:

- Anotar las variants genéticas y las relaciones entre enfermedad-medicamento-genes por medio de la revisión de literatura.
- Sintetizar importantes genes farmacogénicos, asociaciones entre variants genéticas y medicamentos.
- Posibilitar el consorcio examinando importantes cuestiones en relación al conocimiento farmacogénico.
- Participar en la realización de guías escritas sobre las dosis de medicamentos en tratamientos farmacogénicos. Así

mismo, publicarlas.

- Contribuir a la implantación de proyectos farmacogénicos a través de diferentes colaboraciones.
- Difundir la información en su portal Web y permitir descargas [34].

### 2.30. PhysProp

La base de datos de propiedades físicas (denominada PHYSPROP) contiene estructuras químicas, nombres y propiedades físicas de unas 41,000 sustancias químicas. Las propiedades físicas son almacenadas desde una gran variedad de fuentes, e incluyen valores experimentales, extrapolados y estimados sobre el punto de fusión, ebullición, solubilidad del agua, presión de vapor, pKa, etc.

Esta base de datos ha sido construida de forma activa por SRC (Centro de Investigación de Nueva York en temas de seguridad relativa al medio ambiente y la calidad de vida de los ciudadanos) durante las dos últimas décadas.

Empezó siendo una base de datos de propiedades físicas para sustancias químicas que eran evaluadas por SRC para el Banco de Datos de Sustancias denominado HSDB, disponible desde la Biblioteca Nacional de Medicina (NLM). Inicialmente, se evaluaban con cuidado los datos que entraban y eran chequeados por expertos científicos. De forma adicional, se hacía un control de calidad para comparar los valores estimados de diferentes variables.

Esta base de datos se actualiza de forma continua [35].

### 2.31. ProtParam

Se trata de una herramienta que permite la computación de varios parámetros físicos y químicos para una proteína dada almacenada en la base de datos Swiss-Prot<sup>2</sup> o TrEMBL o para una secuencia entera de un usuario. Estos parámetros incluyen el peso molecular, la composición de aminoácidos, la composición atómica, el índice de estabilidad, el tiempo de vida medio estimado, entre otros [36].

### 2.32. PubChem

PubChem es una base de datos de moléculas. Es un sistema operado y mantenido por el National Center for Biotechnology Information (NCBI), parte de la National Library of Medicine, parte a su vez de los National Institutes of Health estadounidenses. Se puede consultar gratuitamente a través de internet, además de poder descargar los datos de millones de estructuras vía FTP. La mayoría de moléculas listadas en PubChem tienen un peso molecular inferior a 2000 uma. La American Chemical Society intentó restringir sus actividades, alegando que competían con su Chemical Abstracts Service. Más de 50 proveedores de bases de datos contribuyen al crecimiento de PubChem [37].

### 2.33. PubMed

PubMed es un motor de búsqueda de libre acceso a la base de datos MEDLINE de citas y resúmenes de artículos de investigación biomédica. Ofrecido por la Biblioteca Nacional de Medicina de los Estados Unidos

como parte de Entrez. MEDLINE tiene alrededor de 4.800 revistas publicadas en Estados Unidos y en más de 70 países de todo el mundo desde 1966 hasta la actualidad [38].

### 2.34. RxList

RxList es un recurso médico on-line destinado a ofrecer información farmacológica detallada sobre medicamentos actuales tanto genéricos como no genéricos. Creada en 1995 por farmacéuticos, RxList es el primer índice de recursos de medicamentos en Internet. Dispone de un diccionario sobre medicamentos con definiciones y abreviaturas [39].

### 2.35. RxNorm

RxNorm suministra nombres normalizados para medicamentos y enlaza sus nombres a vocabularios comúnmente empleados en la gestión de farmacia y en el software de interacción de medicamentos. Es creado por la Biblioteca Nacional de Medicina (NLM) [40].

### 2.36. SIDER

La base de datos SIDER contiene información sobre medicinas de tipo comercial y sus reacciones adversas. La información es extraída de documentos de carácter público. También la información disponible incluye la frecuencia del efecto, medicamento, clasificación del efecto y los enlaces a más información, como por ejemplo las relaciones con otros medicamentos [41].

### 2.37. Spectral Database System (SDBS)

SDBS es una base de datos espectral e integral para componentes orgánicos, que incluye 6 diferentes tipos

dentro de un directorio de components. El número total acumulado de accesos alcanzó los 350 millones a finales de Febrero de 2011. Es de origen japonés [42].

### 2.38. STITCH

STITCH es un recurso para explorar interacciones de productos químicos y proteínas conocidas. Los productos químicos son enlazados a otros y a proteínas derivadas de otros experimentos, bases de datos y precedentes de la literatura. STITCH contiene interacciones para unas 300,000 pequeñas moléculas y 2.6 millones de proteínas procedentes de 1133 organismos [43].

### 2.39. T3DB

La base de datos denominada Toxin and Toxin Target (T3DB) es un recurso bioinformático que combina los datos de toxinas detallados con información de dianas de toxinas. Actualmente está compuesta por 2900 toxinas descritas por más de 34 200 sinónimos, incluyendo pesticidas, medicamentos, toxinas de comida, de las cuales 1300 corresponden a historiales de dianas de toxinas. También está formada por 33 800 toxinas y asociaciones de dianas de toxinas. Cada historial de toxina (ToxCARD) contiene sobre 50 campos de datos y aloja información como es propiedades químicas y descriptores, valores de toxicidad, interacciones moleculares celulares e información de tipo médico. Esta información ha sido extraída de 5600 fuentes, que incluyen otras bases de datos, documentos gubernamentales, libros y literatura científicas. El foco de T3DB está en proveer mecanismos de toxicidad de proteínas de dianas para cada toxina. Esta naturaleza dual de T3DB, en la que toxinas y dianas son enlazadas de forma interactiva en ambas direcciones, la hace única dentro de las bases de datos existentes. Per-

mite también la búsqueda complete de texto, secuencias, estructuras químicas, consultas relacionales. T3DB está alojada en un servidor Apple XServe en el Laboratorio Wishart del campus de la Universidad de Alberta.

rés. Entre ellas: ToxLine, GeneTox, Dart, etc. [47].

### 2.43. UniProt

La base de conocimiento UniProt (denominada Uni-

ProtKB) es el núcleo central de la colección de información funcional sobre proteínas. Se trata de una base de datos consistente, segura y con importantes anotaciones. Para capturar la información obligatoria para cada entrada en UniProtKB (principalmente, secuencias de aminoácidos, nombres o descripciones de proteínas, etc.), es necesario añadir información sobre la anotación. Esto incluye ontologías biológicas que hayan sido aceptadas, clasificaciones, referencias cruzadas, y claras indicaciones de la calidad de la información en el formulario de la

atribución evidente de los datos experimentales y computacionales.

Esta base de conocimiento consta de dos secciones: una sección contiene informes anotados de forma manual con información extraída de literatura y de análisis computacionales realizados, y otra sección consta de informes analizados computacionalmente que están esperando una anotación manual completa. La primera sección se denomina "UniProtKB/Swiss-Prot" (revisada, anotada manualmente) y la otra "UniProtKB/TrEMBL" (no revisada y anotada de forma automática).

Así pues, UniProt es una colaboración entre el Instituto de Bioinformática Europeo (EBI, de las siglas en inglés: European Bioinformatics Institute), el Instituto de Bioinformática Swiss (SIB) y el Centro de Recursos de Información de Proteínas (PIR). Alrededor de 90 personas de los tres institutos trabajan en diferentes

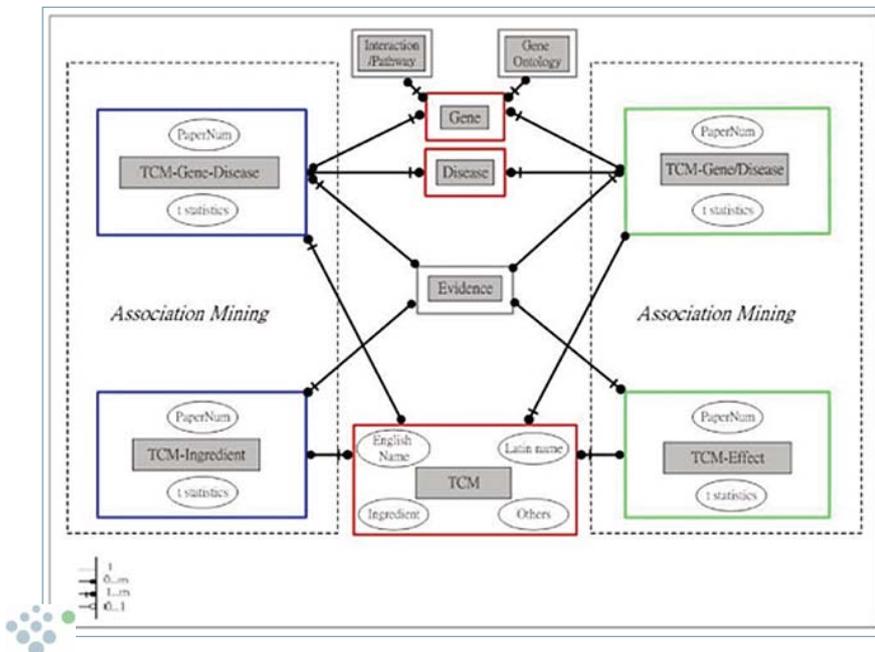


Figura 1. Esquema relacional de la base de datos TCMGeneDIT.

### 2.41. Therapeutic Target Database

Actualmente, esta base de datos contiene 2,025 dianas farmacológicas, incluyendo 364 con éxito, 286 ensayos clínicos, 44 discontinuas y 1,331 de investigación, 7,816 medicamentos, 1,540 aprobadas, 1,423 ensayos, 14,853 experimentales y 3,681 agentes multi-dianas (14,170 pequeñas moléculas y 652 drogas antisensitivas con estructura disponible o secuencia oligonucleótida). Las dianas y los medicamentos en esta base de datos cubren 61 clases de proteínas bioquímicas y 140 clases de medicamentos terapéuticos, respectivamente [46].

### 2.42. ToxNet

Se trata de un conjunto de base de datos sobre toxicología, salud ambiental y productos tóxicos de inte-

tareas como son: mantenimiento de la base de datos, desarrollo y soporte software [48].

### 3. Conclusiones

Existen numerosas fuentes de datos públicas de gran interés para la HCE. La mayoría de ellas son de origen estadounidense y canadiense, también tienen importantes bases de datos Alemania y Japón. El Centro Nacional para la Información Biotecnológica o National Center for Biotechnology Information (NCBI), que es parte de la Biblioteca Nacional de Medicina de Estados Unidos, es uno de los mayores creadores de bases de datos relativas a la salud. Todas sus bases de datos están disponibles en la Web de forma gratuita. NCBI alberga genoma secuenciado en la base de datos GenBank y en artículos biomédicos de investigación en PubMed, así como otra información relevante a la biotecnología. El motor de búsqueda de esas bases de datos es Entrez, al cual se accede a través de: <http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi>.

En este estudio se han analizado 43 de ellas, incluyendo la información más relevante de las mismas, el acceso Web y la apariencia del portal Web para acceder al repositorio digital de las mismas. El portal Web de todas ellas funciona de forma adecuada salvo el de la base de datos TCMGeneDIT. En el Anexo I se puede observar una Tabla resumen de las bases de datos estudiadas.

### Referencias

- [1] Pathak J, Kiefer RC, Chute CG. Applying linked data principles to represent patient's electronic health records at Mayo clinic: a case report, Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium 2012;455-464.
- [2] ADME databases. Disponible en: [http://modem.ucsd.edu/adme/databases/databases\\_extend.htm](http://modem.ucsd.edu/adme/databases/databases_extend.htm). Última visita: 20 de Septiembre de 2012.
- [3] BCCA Cancer Drugs Manual. Disponible en: <http://www.bccancer.bc.ca/HPI/DrugDatabase/DrugIndexPro/default.htm>. Última visita: 20 de Septiembre de 2012.
- [4] CAS Database. Disponible en: <http://www.cas.org/content>. Última visita: 20 de Septiembre de 2012.
- [5] ChEBI Database. Disponible en: <http://www.ebi.ac.uk/chebi>. Última visita: 20 de Septiembre de 2012.
- [6] ChemPDB Database. Disponible en: <http://www.ebi.ac.uk/msd-srv/chempdb>. Última visita: 20 de Septiembre de 2012.
- [7] DailyMed Database. Disponible en: <http://dailymed.nlm.nih.gov/dailymed/mobile/index.cfm>. Última visita: 20 de Septiembre de 2012.
- [8] DBpedia. Disponible en: <http://dbpedia.org/About>. Última visita: 20 de Septiembre de 2012.
- [9] Wheeler DL, Barrett T, Benson DA, et al. Database resources of the National Center for Biotechnology Information, Nucleic Acids Res. 2007, 35 (Database issue): D5-12. DOI:10.1093/nar/gkl1031. PMC 1781113. PMID 17170002.
- [10] Sherry ST, Ward M, and Sirotkin, K. dbSNP - database for single nucleotide polymorphisms and other classes of minor genetic variation, Genome Research 1999, 9(8);677-679. DOI:10.1101/gr.9.8.677. PMID

- 10447503
- [11] dbSNP database. Disponible en: <http://www.ncbi.nlm.nih.gov/projects/SNP>. Última visita: 20 de Septiembre de 2012.
- [12] Diseasome database. Disponible en: <http://diseasome.kobic.re.kr>. Última visita: 20 de Septiembre de 2012.
- [13] Drugbank database. Disponible en: <http://www.drugbank.ca/databases>. Última visita: 20 de Septiembre de 2012.
- [14] Drugs.com. Disponible en: <http://www.drugs.com>. Última visita: 20 de Septiembre de 2012.
- [15] FDA. Disponible en: <http://www.fda.gov/>. Última visita: 20 de Septiembre de 2012.
- [16] Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW, GenBank, Nucleic Acids Res. 2011;39(Database issue):D32-7. Epub 2010 Nov 10.
- [17] Genbank database. Disponible en: <http://www.ncbi.nlm.nih.gov/genbank/>. Última visita: 20 de Septiembre de 2012.
- [18] GeneID database. Disponible en: <http://www.ncbi.nlm.nih.gov/gene/>. Última visita: 20 de Septiembre de 2012.
- [19] Gene Ontology database. Disponible en: <http://www.geneontology.org/>. Última visita: 20 de Septiembre de 2012.
- [20] HapMap database. Disponible en: <http://www.hapmap.org/>. Última visita: 22 de Septiembre de 2012.
- [21] Health Canada DPD. Disponible en: <http://www.hc-sc.gc.ca/dhpm-prodpharma/databasdon/index-eng.php>. Última visita: 22 de Septiembre de 2012.
- [22] HGNC database. Disponible en: <http://www.genenames.org/>. Última visita: 22 de Septiembre de 2012.
- [23] KEGG database. Disponible en: [http://www.freebase.com/view/en/cpd\\_kegg\\_ligand\\_database\\_for\\_chemical\\_compound](http://www.freebase.com/view/en/cpd_kegg_ligand_database_for_chemical_compound). Última visita: 22 de Septiembre de 2012.
- [24] Ligand Depot database. Disponible en: <http://ligand-depot.rutgers.edu/>. Última visita: 22 de Septiembre de 2012.
- [25] LinkedCT. Disponible en: <http://linkedct.org/>. Última visita: 21 de Septiembre de 2012.
- [26] MediCare database. Disponible en: <http://www.cms.gov/medicare-coverage-database/>. Última visita: 22 de Septiembre de 2012.
- [27] NIST database. Disponible en: <http://webbook.nist.gov/chemistry/>. Última visita: 22 de Septiembre de 2012.
- [28] OMIM database. Disponible en: <http://www.ncbi.nlm.nih.gov/omim/>. Última visita: 22 de Septiembre de 2012.
- [29] Open Babel. Disponible en: <http://openbabel.sourceforge.net/>. Última visita: 22 de Septiembre de 2012.
- [30] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, Philip E. Bourne, The Protein Data Bank, Nucleic Acids Research 2000, 28:1; 235-242
- [31] PDB database. Disponible en: <http://www.rcsb.org/pdb>. Última visita: 22 de Septiembre de 2012.
- [32] PDRHealth database. Disponible en: <http://www.pdrhealth.com>. Última visita: 22 de Septiembre de 2012.
- [33] Pfam database. Disponible en: <http://pfam.janelia.org/>. Última visita: 22 de Septiembre de 2012.
- [34] PharmGKB database. Disponible en: <http://www.pharmgkb.org/>. Última visita: 22 de Septiembre de 2012.
- [35] Physprop demo. Disponible en: <http://www.syrres.com/esc/physdemo.htm>. Última visita: 22 de Septiembre de 2012.
- [36] Protparam Server. Disponible en: <http://www.ex->

pasy.org/tools/protparam.html. Última visita: 22 de Septiembre de 2012.

[37] PubChem database. Disponible en: <http://pubchem.ncbi.nlm.nih.gov/>. Última visita: 22 de Septiembre de 2012.

[38] PubMed. Disponible en: <http://www.ncbi.nlm.nih.gov/sites/entrez..> Última visita: 22 de Septiembre de 2012.

[39] RxList database. Disponible en: <http://www.rxlist.com/>. Última visita: 22 de Septiembre de 2012.

[40] RxNorm database. Disponible en: <http://www.nlm.nih.gov/research/umls/rxnorm/>. Última visita: 22 de Septiembre de 2012.

[41] SIDER database. Disponible en: <http://sideeffects.embl.de/>. Última visita: 22 de Septiembre de 2012.

[42] SDBS database. Disponible en: [http://riodb01.ibase.aist.go.jp/sdbs/cgi-bin/cre\\_index.cgi](http://riodb01.ibase.aist.go.jp/sdbs/cgi-bin/cre_index.cgi). Última visita: 22 de Septiembre de 2012.

[43] STITCH database. Disponible en: <http://stitch.embl.de/>. Última visita: 22 de Septiembre de 2012.

[44] T3DB database. Disponible en: <http://www.t3db.org/>. Última visita: 22 de Septiembre de 2012.

[45] Yu-Ching Fang, Hsuan-Cheng Huang, Hsin-Hsi Chen, Hsueh-Fen Juan, TCMGeneDIT: a database for associated traditional Chinese medicine, gene and disease information using text mining, BMC Complementary & Alternative Medicine, 2008, 8:58. Acceso Web: <http://www.biomedcentral.com/1472-6882/8/58/>

[46] TTD database. Disponible en: [http://xin.cz3.nus.edu.sg/group/cjttd/TTD\\_ns.asp](http://xin.cz3.nus.edu.sg/group/cjttd/TTD_ns.asp). Última visita: 22 de Septiembre de 2012.

[47] ToxNet database. Disponible en:

<http://toxnet.nlm.nih.gov/>. Última visita: 22 de Septiembre de 2012.

[48] UniProt database. Disponible en: <http://www.uniprot.org>. Última visita: 22 de Septiembre de 2012.

## ANEXO I: RESUMEN DE FUENTES DE DATOS PÚBLICAS DE INTERÉS EN MEDICINA

Nombre	Web	Creador
ADME	<a href="http://modem.ucsd.edu/adme/databases/databases_extend.htm">http://modem.ucsd.edu/adme/databases/databases_extend.htm</a>	Dr. Tingjun Hou
BCCA Cancer Drugs Manual	<a href="http://www.bccancer.bc.ca/HPI/DrugDatabase/DrugIndex-Pro/default.htm">http://www.bccancer.bc.ca/HPI/DrugDatabase/DrugIndex-Pro/default.htm</a>	Provincial Health Services Authority
CAS	<a href="http://www.cas.org/content">http://www.cas.org/content</a>	American Chemical Society
ChEBI	<a href="http://www.ebi.ac.uk/chebi">http://www.ebi.ac.uk/chebi</a>	European Bioinformatics Institute
ChemPDB	<a href="http://www.ebi.ac.uk/pdbe-srv/pdbechem/">http://www.ebi.ac.uk/pdbe-srv/pdbechem/</a>	European Bioinformatics Institute
DailyMed	<a href="http://dailymed.nlm.nih.gov/dailymed/mobile/index.cfm">http://dailymed.nlm.nih.gov/dailymed/mobile/index.cfm</a>	U.S. National Library of Medicine
Dbpedia	<a href="http://dbpedia.org/About">http://dbpedia.org/About</a>	Open Link Software, Universität Leipzig y Freie Universität Berlin
dbSNP	<a href="http://www.ncbi.nlm.nih.gov/projects/SNP/">http://www.ncbi.nlm.nih.gov/projects/SNP/</a>	National Center for Biotechnology Information
Diseasome	<a href="http://diseasome.kobic.re.kr">http://diseasome.kobic.re.kr</a>	Korean Bioinformation Center
Drugbank	<a href="http://www.drugbank.ca/databases">http://www.drugbank.ca/databases</a>	Genome Alberta & Genome Canada, GenomeQuest, Inc.
Drugs.com	<a href="http://www.drugs.com/">http://www.drugs.com/</a>	
FDA	<a href="http://www.fda.gov">http://www.fda.gov</a>	U.S. Department of Health & Human Services
Genbank	<a href="http://www.ncbi.nlm.nih.gov/genbank">http://www.ncbi.nlm.nih.gov/genbank</a>	National Center for Biotechnology Information
GeneID	<a href="http://www.ncbi.nlm.nih.gov/gene">http://www.ncbi.nlm.nih.gov/gene</a>	National Center for Biotechnology Information
Gene Ontology	<a href="http://www.geneontology.org">http://www.geneontology.org</a>	Berkeley Bioinformatics Open-source Project (BBOP)
HapMap	<a href="http://www.hapmap.org">http://www.hapmap.org</a>	Collaboration among scientists and funding agencies
Health Canada DPD	<a href="http://www.hc-sc.gc.ca/dhp-mps/prodpharma/databas-don/index-eng.php">http://www.hc-sc.gc.ca/dhp-mps/prodpharma/databas-don/index-eng.php</a>	Health Canada
HGNC	<a href="http://www.genenames.org/">http://www.genenames.org/</a>	National Human Genome Research Institute (NHGRI)
KEGG	<a href="http://www.freebase.com/view/en/cpd_kegg_ligand_database_for_chemical_compound">http://www.freebase.com/view/en/cpd_kegg_ligand_database_for_chemical_compound</a>	
Ligan Depot	<a href="http://ligand-depot.rutgers.edu/">http://ligand-depot.rutgers.edu/</a>	Research Collaboratory for Structural Bioinformatics (RCSB)
LinkedCT	<a href="http://linkedct.org/">http://linkedct.org/</a>	
MediCare	<a href="http://www.cms.gov/medicare-coverage-database/">http://www.cms.gov/medicare-coverage-database/</a>	Centers for Medicare & Medicaid Services

NIST	<a href="http://webbook.nist.gov/chemistry/">http://webbook.nist.gov/chemistry/</a>	National Institute of Standards and Technology (NIST), Material Measurement Laboratory (MML)
OMIM	<a href="http://www.ncbi.nlm.nih.gov/omim/">http://www.ncbi.nlm.nih.gov/omim/</a>	National Center for Biotechnology Information
Open Babel	<a href="http://openbabel.org/wiki/Main_Page">http://openbabel.org/wiki/Main_Page</a>	
PDB	<a href="http://www.rcsb.org/pdb/home/home.do">http://www.rcsb.org/pdb/home/home.do</a>	Research Collaboratory for Structural Bioinformatics (RCSB)
PDRHealth	<a href="http://www.pdrhealth.com/">http://www.pdrhealth.com/</a>	
Pfam	<a href="http://pfam.janelia.org/">http://pfam.janelia.org/</a>	Trust Sanger Institute y Howard Hughes Janelia Farm Re- search Campus
PharmGKB	<a href="http://www.pharmgkb.org/">http://www.pharmgkb.org/</a>	NIH/NIGMS. Universidad de Standford
Physprop	<a href="http://www.syrres.com/what-we-do/databaseforms.aspx?id=386">http://www.syrres.com/what-we-do/databaseforms.aspx?id=386</a>	
Protparam	<a href="http://web.expasy.org/protparam/">http://web.expasy.org/protparam/</a>	SIB Swiss Institute of Bioinformatics
PubChem	<a href="http://pubchem.ncbi.nlm.nih.gov/">http://pubchem.ncbi.nlm.nih.gov/</a>	National Center for Biotechnology Information
PubMed	<a href="http://www.ncbi.nlm.nih.gov/sites/entrez">http://www.ncbi.nlm.nih.gov/sites/entrez</a>	National Center for Biotechnology Information
RxList	<a href="http://www.rxlist.com/script/main/hp.asp">http://www.rxlist.com/script/main/hp.asp</a>	RxList Inc.
RxNorm	<a href="http://www.nlm.nih.gov/research/umls/rxnorm/">http://www.nlm.nih.gov/research/umls/rxnorm/</a>	Biblioteca Nacional de Medicina
SIDER	<a href="http://sideeffects.embl.de/">http://sideeffects.embl.de/</a>	Structural and Computational Biology Unit, European Molecular Biology Laboratory
SDBS	<a href="http://riodb01.ibase.aist.go.jp/sdbs/cgi-bin/cre_index.cgi">http://riodb01.ibase.aist.go.jp/sdbs/cgi-bin/cre_index.cgi</a>	National Institute of Advanced Industrial Science and Technology (AIST), Japan.
STITCH	<a href="http://stitch.embl.de">http://stitch.embl.de</a>	
T3DB	<a href="http://www.t3db.org">http://www.t3db.org</a>	Genome Alberta & Genome Canada
TCMGeneDIT	<a href="http://tcm.lifescience.ntu.edu.tw">http://tcm.lifescience.ntu.edu.tw</a>	
TTD	<a href="http://xin.cz3.nus.edu.sg/group/cjttd/TTD_ns.asp">http://xin.cz3.nus.edu.sg/group/cjttd/TTD_ns.asp</a>	Department of Computational Science   National Univer- sity of Singapore
ToxNet	<a href="http://toxnet.nlm.nih.gov">http://toxnet.nlm.nih.gov</a>	Specialized Information Services (SIS) - USA
UniProt	<a href="http://www.uniprot.org">http://www.uniprot.org</a>	European Bioinformatics Institute, SIB Swiss Institute of Bioinformatics, Protein Information Resource (PIR)



RevistaeSalud.com es una publicación electrónica que intenta promover el uso de TICs (Tecnologías de la Información y las Comunicaciones) con el propósito de mejorar o mantener la salud de las personas, sin importar quiénes sean o dónde estén.

Edita: FESALUD – Fundación para la eSalud  
Correo-e: [cperez@fesalud.org](mailto:cperez@fesalud.org)  
ISSN 1698-7969

Los textos publicados en esta revista, a menos que se indique lo contrario, están sujetos a una licencia de Reconocimiento-NoComercial-SinObraderivada 2.5 de Creative Commons. Pueden copiarse, distribuirse y comunicarse públicamente, siempre que se citen el autor y la revista digital donde se publican, RevistaeSalud.com. No se permite su uso comercial ni la generación de obras derivadas. Puede consultarse la licencia completa en:  
<http://creativecommons.org/licenses/by-nc-nd/2.5/deed.es>