

James J. Cimino¹

¹Chief of the Laboratory for Informatics Development at the Clinical Center of the US National Institutes of Health. Senior Investigator, US National Library of Medicine. Adjunct Professor of Biomedical Informatics at Columbia University and Georgetown University (EE.UU.)

Retos para construir y utilizar una base de datos para la Investigación clínica

Reptes per construir i emprar una base de dades per Recerca clínica

Challenges to build and use a Clinical Research Data Warehouse

Resumen

Resumen de la conferencia que impartió el Prof. James J. Cimino, durante la X Jornada del Fòrum CIS, que se celebró el pasado mes de abril en Barcelona.

La RevistaeSalud.com es el medio oficial de esta jornada, contando con un especial en su número 31, correspondiente al mes de julio de 2012. Bajo el lema "Envers una assisència sostenible centrada en el pacient", este foro ha reflexionado sobre prácticas y experiencias sanitarias basadas las TIC. En este especial, se ofrecen los resúmenes de las conferencias, mesas redondas y comunicaciones de este evento, en tres idiomas: castellano, catalán e inglés.

Resumen (castellano)

El Prof. Cimino lleva 4 años trabajando en el NIH, que está formado por 27 instituciones diferentes en Bethesda, al norte de Washington DC. Su trabajo consiste en crear un gran repositorio de información clínica orientado hacia la utilización per la investigación a nivel estatal. Los datos más antiguos corresponden al año 1976. Una de las principales características en el desarrollo del proyecto es la heterogeneidad de formatos y soportes de la información clínica de interés per constituir el repositorio. Aunque los datos son los mismos, la visión es diferente si quien los observa es un clínico o es un investigador. El proyecto que desarrolla el Prof. Cimino se conoce con el nombre de BTRIS (Biomedical

Translational Research Information System) y permite construir un repositorio único que sea capaz de ofrecer distintas vistas que den respuesta a los usuarios independientemente de su profesión. Muestra cifras cuantitativas que demuestran el incremento progresivo de los datos (millones de registros). El sistema está diseñado para que sea autogestionado por el propio investigador con la ayuda de una serie de protocolos predefinidos. Uno de los principales problemas es la terminología. Para resolver este problema se ha desarrollado un servidor de terminología con un diccionario de entidades que unifica los diferentes lenguajes y los distintos significados del concepto en una estructura de jerarquías y relaciones. Es posible explorar la jerarquía por parte del usuario. Finalmente se ejecuta la consulta i se obtiene un listado con los datos solicitados que se pueden importar a diferentes aplicaciones. La misma herramienta también puede analizar documentos clínicos y encontrarlos (completos o en fragmentos) relacionados con palabras clave. Es po-

sible también encontrar imágenes y perfiles de medicación. A mes a mes la herramienta permite graficar los elementos de información en base a la línea del tiempo. Los retos principales han consistido en facilitar la búsqueda de datos concretos relacionadas con



Momento de la conferencia del Dr. Cimino en la X Jornada FòrumCIS.

las fechas. En algunos casos se tuvo que recurrir a un museo informático para poder leer las cintas magnéticas con sistemas que ya no existen en la actualidad. Otro problema importante son los metadatos necesarios per un registre completo. A veces porque no son suficientemente explícitos, y otras veces porque coexisten diferentes opciones. Presenta un par de ejemplos que ponen en evidencia la utilidad del sistema BTRIS per descubrir información a partir del análisis de los datos. En estos momentos empiezan a almacenar datos genéticos. Los principales puntos en los que está trabajando son: mejorar todavía más el interface de usuario para ampliar el espectro de profesionales que puedan realizar consultas; facilitar al máximo la comprensión de la estructura y de la terminología utilizada y facilitar la realización de búsquedas complejas. Entre los beneficios existentes está la accesibilidad a los datos, la existencia de los protocolos de extracción que permiten la ejecución inmediata, la agregación de casos por criterios comunes, la capacitación del usua-

rio y las búsquedas basadas en conceptos. BTRIS se ha convertido en un elemento atractivo por su valor añadido y ahora son los propios profesionales que quieren volcar sus datos y mediante sus comentarios es posible marcar las prioridades de los desarrollos futuros. BTRIS puede consultarse en <http://btris.nih.gov>. En la discusión se le pide opinión sobre las versiones 10 i 11 del ICD y su valor para la investigación y responde que en su artículo titulado "Desiderata" explica los grandes



Se ha desarrollado un servidor de terminología con un diccionario de entidades que unifica los diferentes lenguajes y los distintos significados del concepto en una estructura de jerarquías y relaciones



defectos que tiene ICD. La razón es tria en que ICD es un simple catálogo y no contempla es aspecto conceptual y las relaciones entre las entidades. Además, los médicos y los clínicos no tie ni habilidades suficientes para codificar. Es mejor dejar esta tarea a los documentalistas para satisfacer necesidades como la elaboración de estadísticas o la facturación. Existen terminologías mucho mejores para utilizar en clínica e investigación como es el cas de SNOMED. También se le pregunta sobre el valor económico que puede representar una solución como BTRIS en un mercado abierto. Cimino responde que la pregunta es difícil de contestar y que en el cas de BTRIS era algo que el NIH consideraba que debía hacerse necesariamente por los beneficios que comportaría y no por que lo indicara un cálculo del ROI.

Resum (català)

El Prof. Cimino porta 4 anys treballant al NIH, que està format per 27 institucions diferents a Bethesda, al nord de Washington DC. El seu treball consisteix en crear un gran repositori d'informació clínica orientat en la seva utilització per la recerca a nivell estatal. Les dades introduïdes més antigues son de l'any 1976. Una de les principals característiques en el desenvolupament del projecte es la heterogeneïtat de formats i suports de la informació clínica d'interès per incloure al repositori. Tot i que les dades son les mateixes, la visió de les dades es diferent si qui els mira es un clínic o es un investigador. El projecte que desenvolupa el Prof. Cimino es anomenat BTRIS (Biomedical Translational Research Information System) i permet construir un repositori únic que sigui capaç de oferir diferents visions que donin resposta als usuaris independentment del seva professió. Mostra dades quantitatives que demostren el increment progressiu de les dades (milions de registres). El sistema està dissenyat perquè sigui autogestionat pel propi investigador amb l'ajuda d'una colla de protocols predefinits. Un dels problemes principals es la terminologia. Per resoldre aquest problema s'ha hagut de desenvolupar un servidor de terminologia amb un diccionari d'entitats que unifica els diferents llenguatges i els diferents significats del concepte en una estructura de jerarquies i relacions. Es possible explorar la jerarquia per part del usuari. Finalment s'executa la consulta i s'obté un report amb les dades sol·licitades que es poden importar a diferents aplicatius. La mateixa eina també te capacitat d'analitzar documents clínics i trobar-los (complets o en fragments) relacionats amb paraules clau. Es possible també trobar imatges i perfils de medicació. A mes a mes la eina permet fer gràfiques amb els elements d'informació en base a la línia del temps. Els reptes principals ha estat facilitar el trobar les dades relacio-

nades amb les dates. En alguns casos es va tenir de portar cintes magnètiques a un museu informàtic per poder-les llegir amb sistemes que ja no existeixen. Altre problema important son les metadades necessàries per un registre complet. De vegades perquè no son prou explícites, i altres cops perquè hi ha diferents opcions. Presenta un parell d'exemples que posen en evidència la utilitat del sistema BTRIS per descobrir informació a partir de la anàlisi de les dades. En aquests moments s'estan començant a emmagatzemar les dades genètiques. Els punts principals en els que està treballant son: millorar encara mes el interfacie de l'usuari per ampliar l'espectre de professionals que puguin fer interrogacions; facilitar al màxim la comprensió de la estructura y de la terminologia emprada i facilitar la realització de cerques complexes. Entre els beneficis existents s'hi troba l'accessibilitat a les dades, l'existència dels protocols d'extracció que permeten l'execució immediata, l'agregació de casos per criteris comuns, la capacitat del usuari i les cerques basades en conceptes. BTRIS s'ha convertit en un element atractiu pel seu valor afegit i ara son els propis professionals que hi volen abocar les seves dades i els comentaris dels usuaris serveixen per marcar les prioritats dels desenvolupaments futurs. BTRIS es accessible a <http://btris.nih.gov>. En la discussió se li planteja la opinió que te sobre les versions 10 i 11 del ICD i el seu valor per la recerca i respon que al seu article anomenat "Desiderata" explica els grans defectes que te la ICD. La raó es que ICD es un simple catàleg i no contempla els conceptes i las relacions entre elles. A mes a mes els metges i clínics no tenen temps i tam-



Millorar encara mes el interfacie de l'usuari per ampliar l'espectre de professionals que puguin fer interrogacions; facilitar al màxim la comprensió de la estructura y de la terminologia emprada i facilitar la realització de cerques complexes



poc habilitats per codificar. Es millor deixar-ho als documentalistes per les utilitats que te que son estadística o facturació. Hi han terminologies molt millors per emprar en clínica i recerca com es el cas de SNOMED.

També se li pregunta sobre el valor econòmic que pot representar una solució com BTRIS en un mercat obert. Cimino respon que la pregunta es difícil de respondre i que en el cas de BTRIS era quelcom que el NIH

considerava que s'havia forçosament de fer pels beneficis que comportaria i no per que un càlcul del ROI ho indiqués.

Summary (English)

The speaker begins with a description about the "macro" scenarios of different models of healthcare organizations and their funding which are going to condition in large measure the "micro" scenarios focused on the clinical management of patients what in fact, is really relevant. The interrelationship between macro and micro is total, being decisive the grade co-help in this binomial to make easier the integral and continuum of care. A fundamental difference is that "macro" settings are based in the citizen, whereas "micros" are centred in the patient and in the middle are the clinic professionals and the healthcare managers. Rather than the "sustainability" of the sanitary system, which is an aspect meditate by the political decisions, speaker concerns are more related to the reliability as the resolution capacity against changing situations:

aging, increasing demand, etc., Another important concept is that "more sometimes is not best". "More" means more population covered by the system, but the reliability depends in one hand about the possibility to take on the waste generate by the redistributive capacity that consists in the decongestion of the public system thanks to the high income rents that consume their own healthcare in alternative systems (private medicine, insurances, benefit societies, etc) and by the other side, specially in the field of investment, depends on the incomes by taxes which are based on indirect taxes and labour rents. With this approximation it's possible to begin a regressive drift which encourages the proposal of materialize the "co-payment" to avoid a reduction of the services. Another idea is the integration of the value chain of healthcare production, including drugs. The Public Healthcare System should

have a global surrounding financial layer which allows a periodic evaluation of the system to know really what it's done and not what is said, and have a real knowledge about the cost-effectiveness, identifying and promoting the quality.

At the question of the audience about the role of technologies the answer is that it depends of the professionals who are the persons that are going to use it to solve the problems. A central imposition (top-down) from the Funding Authority can be seen as more efficient, but it's in a high risk to be not accepted and thereby discarded.

At the question about "what can incentive organizations like those proposed" the answer is that "the need will create the virtue" when the professionals understand the magnitude of the problematic that we are talking. It wouldn't be more remedy than move to

the proposed way to sustain a reliable system

At the question "if it there are still at time to save the situation" the answer is affirmative, specially keeping an open mind and forgetting euphemisms that don't help at all to found and answers to this problem.

At the question "If the actuaciones proposed in Catalonia are going to be accepted by other communities"



El Dr. Cimino durante la conferencia.

the answer is more pessimist because the experience says that in the models very split and heterogeneous like the Spanish one, the maximum common denominator to achieve the consensus is really low.

Videostreaming:

http://www.youtube.com/watch?feature=player_detailpage&v=QsMkeanPMUI

Presentación:

<http://www.forumcis.cat/LinkClick.aspx?fileticket=mujkYUZ9YnE%3d&tabid=170>



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
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>