Diseño y generación de un repositorio de información de salud para ciencia de datos

Design and creation of a repository of health information for data science

María Luisa Ávila Morales[a], Víctor Morales Rocha[a*]

[a]Departamento de Ingeniería Eléctrica y Computación, Maestría en Cómputo Aplicado, Universidad Autónoma de Ciudad Juárez, México

*Autor de correspondencia. Correo: victor.morales@uacj.mx

No. de resumen
2CP21-121

Formato
Ponencia

Evento
2.º Coloquio de Posgrados del IIT

Presentador
María Luisa Ávila Morales

Tema
Cómputo Aplicado

Estatus
Estudio en curso

Fecha de la presentación
Noviembre 11, 2021

Resumen

El objetivo de este proyecto es el diseño y generación de un repositorio de información de salud, disponible bajo una licencia de datos abiertos con el propósito de realizar ciencia de datos. Considerando la cantidad de información que se manejará se utilizan herramientas de Big Data para la creación del repositorio, la creación de un modelo de extracción de datos y de un modelo de procesamiento y carga de información. La información que contendrá el repositorio proviene de diversos sistemas heterogéneos de servicios de salud públicos y privados de México, los cuales se extraerán con la plataforma de interoperabilidad que se está desarrollando en una fase precedente a este proyecto. La información que contendrá el repositorio ya ha sido seleccionada del conjunto de datos clínicos que conforman el Expediente Clínico Electrónico (ECE) según las guías de intercambio de información de la Secretaría de Salud. Durante el diseño de la solución, fase aún en proceso, se identificaron tecnologías que impactan técnicamente este proyecto. Dada la complejidad de la información se trabaja con una plataforma de base de datos NoSQL en la nube para la creación de un prototipo. El repositorio se publicará para su análisis y explotación, el acceso a este será controlado por medio de un registro de usuario. Se espera que el repositorio de salud facilite la investigación clínica y epidemiológica, así como a la planeación adecuada de servicios de salud, emisión de políticas públicas, lineamientos y normas basadas en información real.

Palabras clave: expediente clínico electrónico; datos clínicos; servicios de salud; ciencia de datos; *Big Data*.

Abstract

The objective of this project is the design and creation of a repository of health information for data science purpose under open data license. Taking into consideration the amount of information involved, Big Data tools are being used for the creation of the repository, the creation of a data extraction model and the creation of a data process and data load model. This repository will hold information that comes from various heterogeneous systems of public and private health services in Mexico, which will be extracted with the interoperability platform that is being developed in a phase preceding this project. The information the repository will contain has already been selected from the set of clinical data from the Electronic Health Record (EHR) based on the data information exchange guidelines from the Secretariat of Health. During the design phase of the project, still in progress, some technologies were identified to have technical impact to the project. Given the complexity of the information, work has been initiated on a NoSQL database platform in the cloud for the creation of a prototype. The health information will be published for its analysis and exploitation, its access will be controlled by a user registration request. Is to be expected the health data repository to facilitate clinical research – mainly epidemiologic research- as well as an adequate health service planification, the emission of public politics, guides lines and norms based on real information.

Keywords: electronic health record; clinical data; health care services; data science; Big Data.

Entidad legal responsable del estudio

Universidad Autónoma de Ciudad Juárez.

Financiamiento

Los autores. Beca CONACYT del autor principal.

Conflictos de interés

Ninguno.