

Implementación de una arquitectura Open Health Data para la gestión y consumo de datos de salud

Walter Hugo Arboleda Mazo, Jeovany Rojas Medina y Omer Urango
Grupo de Investigación en Ingeniería Aplicada
Corporación Universitaria Adventista

Resumen. En el presente proyecto de investigación se realiza la implementación de una arquitectura Open Health Data para la gestión y consumo de datos de salud. Después de analizar las diferentes soluciones informáticas para la ejecución de este proyecto se optó por utilizar una arquitectura basada en un servidor con sistema operativo Linux y CKAN como software para la gestión y consumo de datos. Los datos allí publicados principalmente obedecen al Programa de Promoción y Prevención de Enfermedades Crónicas no Transmisibles para el departamento de Antioquia, acorde a los lineamientos establecidos por la OMS y la OPS, proyecto perteneciente al Grupo de Investigación de la Facultad de Ingeniería de la Corporación Universitaria Adventista.

Índice de términos: Open Health Data, Arquitectura, e-Health, salud, datos abiertos, metadatos.

I. INTRODUCCIÓN

Según el Ministerio de Salud, las enfermedades no transmisibles representan la causa de defunción más importante del mundo, con un 63% del número total de muertes anuales [1]. La Organización Mundial de la Salud está aplicando todos los esfuerzos para reducir el impacto de las ENT en las personas, por medio de la promoción de hábitos saludables [2]. Por otro lado, la Organización Panamericana de la Salud ya tiene su plan de acción para implementar en las Américas, para la prevención y control de enfermedades no transmisibles, enfocado en promover políticas de salud pública, apoyándose en actores interesados en colaborar. Dados todos estos esfuerzos que realizan diferentes organizaciones de salud, surge la necesidad de crear un programa para la promoción y prevención de enfermedades no transmisibles y que pueda ser utilizado abiertamente por la población.

Es por esta razón por lo que se decide crear una arquitectura de Open Health Data para la gestión y consumo de datos de salud, donde de manera abierta, libre y sin ninguna restricción se pueda acceder a los datos.

La solución se compone de un catálogo de datos abiertos en salud utilizando CKAN, software de código abierto, libre, gratuito y utilizado por muchas otras iniciativas a nivel mundial. Este software se encuentra instalado sobre un servidor con sistema operativo Linux, que también es de código abierto, libre y gratuito.

En este catálogo se publica todo el programa de promoción y prevención de enfermedades no transmisibles, donde puede ser ofrecido sin ninguna reserva, de forma libre y sin restricciones para que sea consumido por la población.

II. DESARROLLO

A. Estado del arte

Un primer objetivo de nuestro proyecto fue hacer una búsqueda sobre catálogos de salud ya existentes. Se pudo encontrar que Colombia y Uruguay son los primeros países de América con mejores iniciativas de datos abiertos en Latinoamérica [3]. En Colombia, el Estado a través del MinTIC como parte de la estrategia de gobierno en línea está obligando a las entidades públicas a publicar sus datos que puedan ser usados para beneficio de la comunidad [4]. Actualmente, tiene su catálogo de datos abiertos, donde más de 500 entidades han publicado más de 2.000 conjuntos de datos. Al analizar específicamente la categoría de la salud, se pudo encontrar que solo con-

tiene 179 conjuntos de datos a la fecha (mayo de 2016) y no existe información acerca de promoción y prevención de enfermedades no transmisibles, lo que es un indicador de que aún falta mucho por hacer. Y es una prueba de que este proyecto es de vital relevancia para el departamento de Antioquia y para el país. Estamos siendo pioneros en la publicación de datos abiertos específicamente para la promoción y prevención de enfermedades no transmisibles en Colombia.

Además del Estado colombiano existen otras entidades públicas y privadas que tienen su catálogo de datos abiertos; algunas de ellas son: Zoom Online, FONADE, Alcaldía de Tuluá, Alcaldía de Bogotá, Alcaldía de Cali, etc. Cada una tiene unos intereses específicos en la publicación de los datos, pero ninguna contiene datos de salud relacionados con la promoción y prevención de ENT.

A nivel mundial se pueden encontrar catálogos de datos abiertos muy completos como son: Data.gov de Estados Unidos, y Data.gov.uk del Reino Unido; también se pueden encontrar otros catálogos como el Banco Mundial, Dbpedia, Estado de México, Uruguay, etc.

En cuanto a la publicación específica de datos de salud, se encuentra healthdata.gov de Estados Unidos con información de alto valor para empresarios, investigadores y políticos, con el objetivo de mejorar la salud de todos [5]. También se pueden encontrar otros catálogos de datos abiertos de salud como Health Datapalooza, Health Data Consortium, Open Health Data Platform, A tu servicio, Journal open

health data, eHealth Ireland, etc.

Según la Fundación CTI, en información actualizada a 19/04/2016, existe un total de 289 catálogos de datos abiertos a nivel mundial [6].

B. Caracterización del software para la arquitectura Open Health Data

1) Caracterización y clasificación del Sistema operativo:

Se hace un estudio y se evalúan las diferentes características de Linux, Windows y MacOS y se decide utilizar en este proyecto el sistema operativo Linux, porque posee características especiales que lo diferencian de los otros sistemas operativos, como la potencia, estabilidad, modificabilidad, portabilidad y, lo más importante, es gratis; todas estas bondades lo hacen el sistema operativo perfecto, además que actualmente está posicionado como líder en el ámbito de internet [7].

Todos los estudios realizados coinciden en que los costes de hardware y software son mucho más económicos en una solución Linux que una basada en otro sistema operativo como Windows o MacOS. En el caso del software Linux, se puede descargar de manera gratuita sin ninguna licencia por la que se deba pagar [8].

Es así como se instala un servidor en Linux, específicamente con la versión de Ubuntu Server 12.04 con los siguientes requerimientos de hardware:

- * 8 GB de memoria RAM
- * 16 GB de espacio en disco
- * Procesador de doble núcleo
- * Tarjeta de red gigabit

Una vez se instala Linux debe cumplir los siguientes requerimientos de sistema:

- * Python 2.6
- * PostgreSQL 9.3
- * Jetty
- * OpenJDK 6
- * Apache
- * Nginx
- * Tomcat, embarcadero + Solr

2) Caracterización y clasificación del software para el catálogo de datos abiertos:

Se hace una búsqueda de las diferentes soluciones de software que se pueden utilizar para la implementación de un catálogo de datos abiertos, se pueden encontrar varias como CKAN, Ogoov, Socrata Open Data Portal, Junar, DKAN, Datapress, OpenDataSoft, OGPL, OGDI, entre otras.

El software que clasificó para nuestro proyecto es CKAN, básicamente por las siguientes cuatro grandes razones: Primero, porque es de código abierto, libre, gratuito y permite una sencilla y rápida implementación. Segundo, porque cuenta con una interfaz amigable y fácil de entender, lo que favorece la navegación del sitio. Tercero, porque CKAN ha sido elegido por muchas otras iniciativas a nivel mundial, lo que prueba la efectividad y eficiencia de la plataforma. Y por último, porque posee una fuerte comunidad activa de usuarios y desarrolladores que impulsan constantemente el avance de esta tecnología [9].

Sobre el servidor ya configurado con Linux se hace la instalación de CKAN, donde una vez instalado se podrá acceder a una interfaz gráfica por medio de un navegador, en el cual se hacen la publicación y el consumo de los conjuntos de datos.

C. Diseño de la arquitectura Open Health Data

La arquitectura está compuesta por un servidor web en Linux, allí se encuentra alojado el software CKAN, utiliza un motor de base de datos que es libre también como es PostgreSQL, tiene su entorno de desarrollo por consola. A la plataforma se puede acceder desde cualquier parte por medio de un navegador web a una url específica, donde se podrá desde un entorno gráfico gestionar la importación de los conjuntos de datos al catálogo, gestión de usuarios con los permisos que sean pertinentes y el consumo de datos.

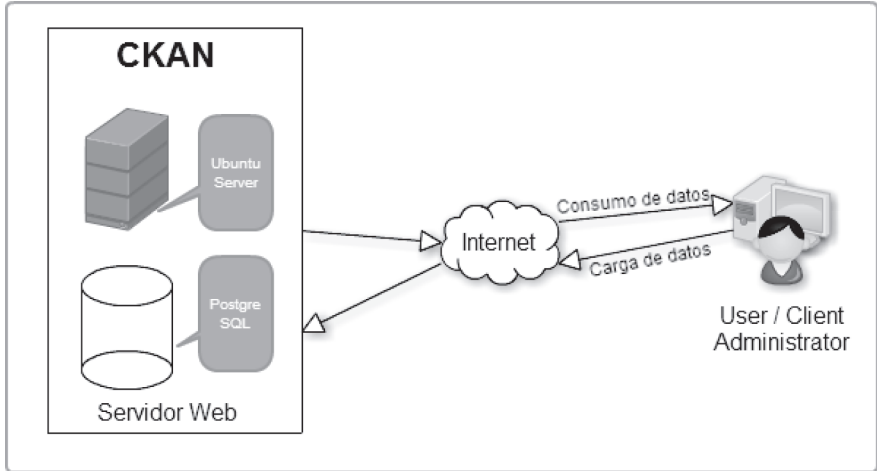


Figura 1: Arquitectura del proyecto.

1) Arquitectura del proyecto del Grupo de Investigación de la Facultad de Ingeniería de la UNAC:

A continuación se presenta la arquitectura general del proyecto que adelanta el Grupo de Investigación de la Facultad de Ingeniería de la UNAC sobre el “Programa de Promoción y Prevención de Enfermedades Crónicas no Transmisibles para el departamento de Antioquia, acorde a los lineamientos establecidos por la OMS y la OPS”.

Se ha desarrollado una aplicación móvil que promueve una guía personal de ocho semanas, que pretende que los usuarios adopten hábitos saludables que los lleven a tener un buen estilo de vida, previniendo las enfermeda-

des no transmisibles; cada semana se les introduce un hábito o una temática nueva, proponiéndoles actividades por desarrollar durante el día en que deberán ir registrando en la aplicación.

Al usar la aplicación por primera vez, el usuario deberá registrar unos datos previamente. La aplicación móvil se comunica con la aplicación web y le envía esos datos. La aplicación web los recibe y los almacena en una base de datos, donde también se pueden visualizar por medio de la aplicación.

Una vez se pone en marcha la plataforma de CKAN se debe hacer la integración entre la aplicación web y la plataforma; mediante una API se logra consumir los datos de la aplicación

web desde la plataforma de CKAN, donde serán publicados automáticamente en el catálogo de datos abiertos. Es conveniente resaltar que por derechos de habeas data no se publicarán datos personales de los participantes. Solo información de interés general para estudios estadísticos.

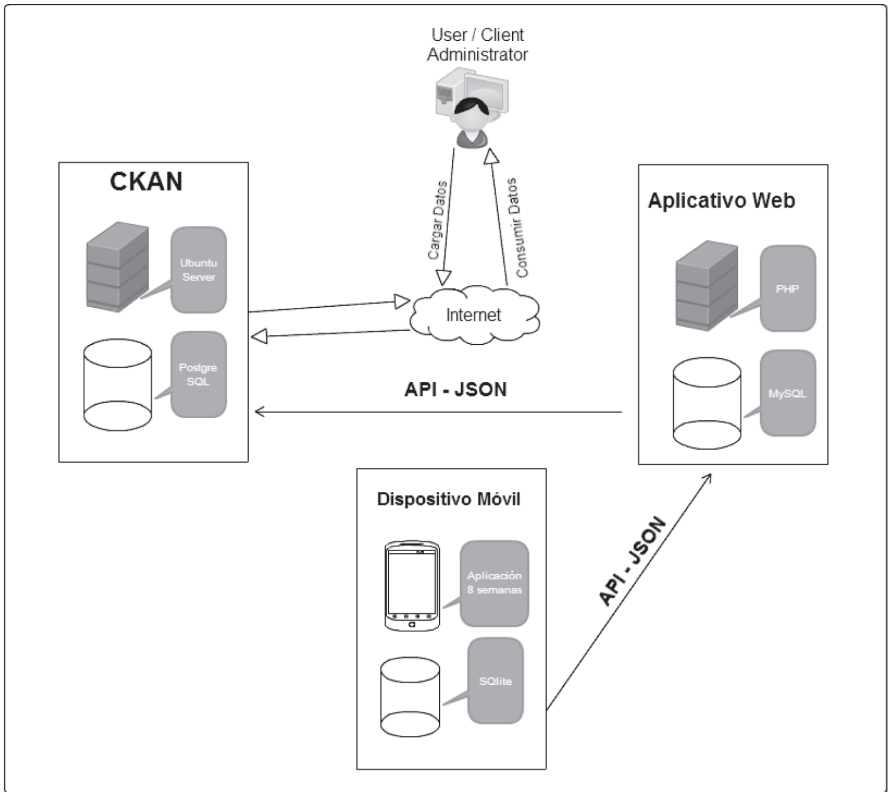


Figura 2: Arquitectura del Proyecto del Grupo de Investigación de la Facultad de Ingeniería de la UNAC.

D. Lineamientos para la publicación de datos

El MinTIC propone la siguiente metodología para la apertura de datos abiertos [10]:



Figura 3: Metodología de apertura de datos.

1) Identificación de la información:

En esta fase se analiza toda la información existente y se caracteriza la que cumple con los requisitos para convertirse en datos abiertos.

2) Análisis de información publicable:

En esta fase se hace la clasificación de la información identificada en la primera fase y se determinan la que puede llegar a publicarse y la que no, de acuerdo con la normativa colombiana.

3) Priorización de los datos:

En esta fase se definen los parámetros que permitirán priorizar los datos que serán publicados en el catálogo. Se medirán el impacto de los datos y la dificultad de implantación.

4) Documentación:

En esta fase se hace la descripción de los datos por publicar para que sean fáciles de entender, buscar y usar por los beneficiarios y consumidores de los datos del catálogo. Para la optimización de búsqueda, en esta fase se hace la descripción de los metadatos.

La ficha de descripción está compuesta por los siguientes tipos de metadatos:

- * Metadatos comunes: Son aquellos que todos los datos los tienen, sin importar su naturaleza; son metadatos generales que se aplican a cualquier dato.
- * Metadatos específicos: Son aquellos que se usan de acuerdo con la naturaleza del dato, son particulares y específicos para cada conjunto de datos.

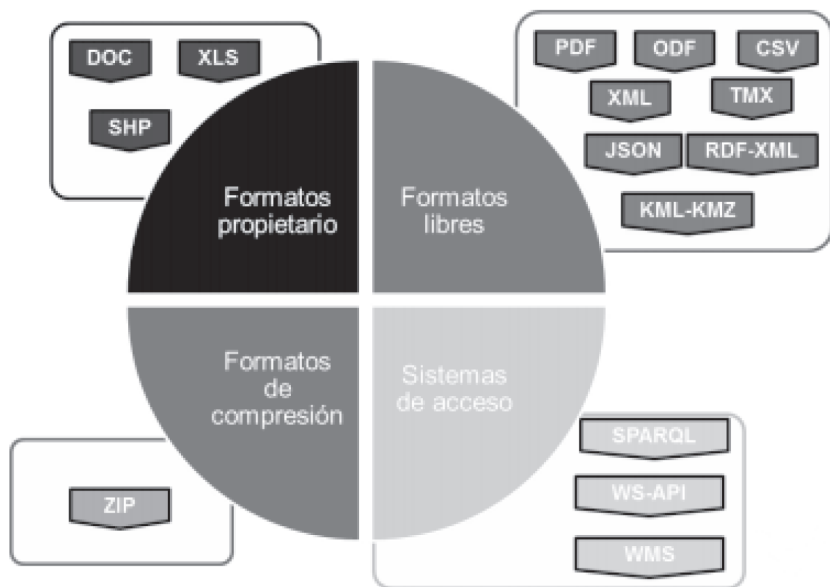


Figura 4: Formato de conjuntos de datos.

- * Formatos propietarios: Estos formatos requieren herramientas que no son públicas para la visualización o edición de la información.
- * Formatos libres: Son formatos de archivos que se pueden manipular para cualquier software, libre de restricciones.
- * Sistemas de acceso: Son lenguajes para la consulta de archivos.
- * Formatos de comprensión: Son formatos que comprimen la información a través de algoritmos.

5) Estructuración, cargue y publicación:

En esta fase se definen los lineamientos para el cargue y la publicación de datos en la plataforma de datos abiertos, donde podrán ser consumidos por los ciudadanos y sociedad en general.

Para la publicación de datos abiertos se debe tener en cuenta la siguiente normativa [11]:

- * Ley 23 de 1981: Determina que tienen carácter de reservadas, las historias clínicas, estableciendo como necesaria la autorización del paciente para que dicha información sea otorgada a un tercero.
- * Ley 95 de 1985: Determina que tienen carácter de reservados, ciertos documentos que reposen en la Registraduría Nacional del Estado Civil, relacionados con temas como los de la identidad de las personas, datos biográficos, entre otros.
- * Ley 30 de 1986: Determina como reservados, datos referentes

al Consejo Nacional de Estupefacientes.

- * Decreto 624 de 1989: Determina el carácter de reservado de cierta información que aparezca en las declaraciones tributarias.
- * Ley 43 de 1990: Otorga el carácter de reservada a la información suministrada al contador, y se aplica la figura del secreto profesional.
- * Ley 52 de 1990: Establece que tiene carácter de reservada cierta información del Consejo Nacional de Seguridad.
- * Decreto 300 de 1997: Determina el carácter de reservado de cierta información del Instituto Nacional Penitenciario y Carcelario (INPEC).
- * Ley 766 de 2002: Por medio de la cual se aprueba la “Convención sobre asistencia en caso de accidente nuclear o emergencia radiológica”, se protege la confidencialidad de los datos en caso de emergencia nuclear o radiológica.
- * Ley 1266 de 2008: Da un tratamiento especial a aquellos datos de información personal que no tengan el carácter de públicos y que se encuentre almacenada en bases de datos, para preservar el derecho Constitucional al Habeas Data.

E. Integración del proyecto con el proyecto general del Grupo de Investigación de la Facultad de Ingeniería.

Al integrarse la arquitectura de Open Health Data con el proyecto de promoción y prevención de enfermedades no transmisibles podemos ver todo su funcionamiento en el siguiente modelo:

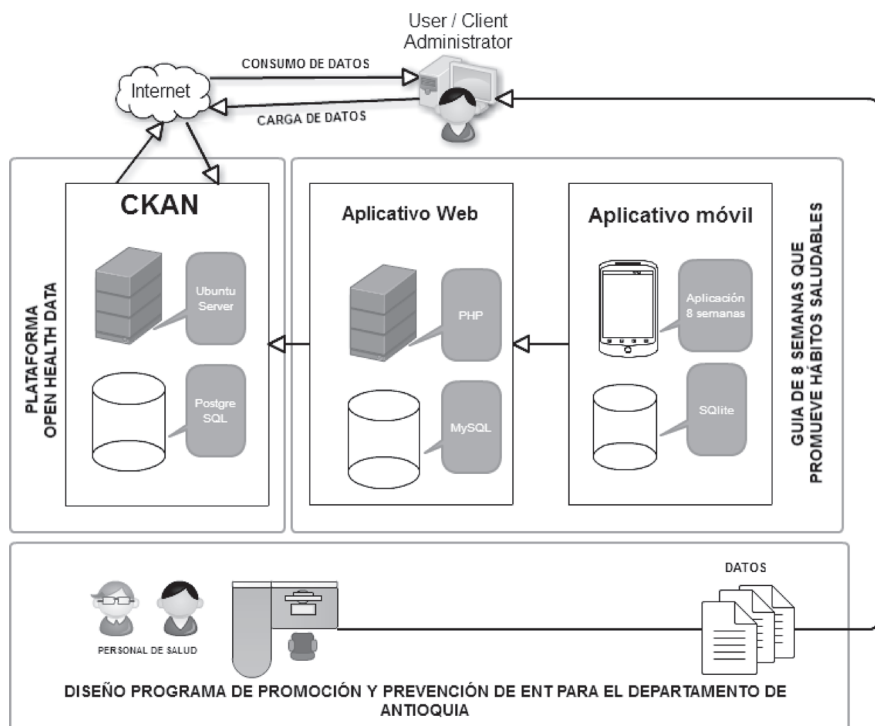


Figura 5: Modelo Open Health Data.

III. CONCLUSIONES

Los resultados de este proyecto de investigación permiten concluir que a la hora de implementar una arquitectura de datos abiertos es recomendable montarla en un servidor con sistema operativo Linux, gracias a su potencia, estabilidad, modificabilidad, portabilidad y, lo más importante, es gratis; además, de que está posicionado como el mejor para soluciones en el ámbito de internet.

Por otro lado, al seleccionar el software para el catálogo de datos abiertos es muy importante elegir uno que sea de código abierto, libre, gratuito, sen-

cillo, fácil de entender y que ya haya sido utilizado por otras iniciativas; estas características las cumple bien CKAN, software que ha sido utilizado para nuestro proyecto.

Al revisar el estado del arte se pudo concluir que nuestro proyecto es muy relevante y además pertinente al pretender compartir datos relacionados con la promoción y prevención de enfermedades no transmisibles, ya que aunque las organizaciones de salud están haciendo esfuerzos por implementar planes de acción, aún no se ven reflejados en los catálogos abiertos de nuestro país.

REFERENCIAS

- [1] Ministerio de Salud, “Prevención de enfermedades no transmisibles.” [Online]. Available: <https://www.minsalud.gov.co/salud/Paginas/Enfermedades-no-transmisibles.aspx>. [Accessed: 08-Apr-2016].
- [2] OMS, “OMS | Enfermedades no transmisibles.” [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs355/es/>. [Accessed: 08-Apr-2016].
- [3] “Datos abiertos: ya hay 57 ‘apps’ hechas con datos abiertos del Estado - Novedades tecnología - ELTIEMPO.COM.” [Online]. Available: <http://www.eltiempo.com/tecnosfera/novedades-tecnologia/datos-abiertos-ya-hay-57-apps-hechas-con-datos-abiertos-del-estado/16157736>. [Accessed: 27-Oct-2015].
- [4] “Catálogo de Datos del Estado Colombiano.” [Online]. Available: <http://www.datos.gov.co/frm/Acerca/frmAcercaDe.aspx>. [Accessed: 28-Jul-2015].
- [5] “HealthData.gov.” [Online]. Available: <http://www.healthdata.gov/content/about>. [Accessed: 06-Oct-2015].
- [6] “Open Data @ CTIC » Sandbox » - Public Dataset Catalogs Faceted Browser {selecciona Tabla vista}.” [Online]. Available: <http://datos.fundacionctic.org/sandbox/catalog/faceted/>. [Accessed: 19-Apr-2016].
- [7] Ciberaula, “Qué es Linux ? - Sistema Operativo Linux - Ciberaula.”
- [8] F. Miralles and G. Armelini, Linux y software de código abierto: ¿listos para su empresa?, Deusto - P. España, 2004.
- [9] “Acerca de Buenos Aires Data.” [Online]. Available: <http://data.buenosaires.gob.ar/about>. [Accessed: 15-Mar-2016].
- [10] “GUÍA PARA LA APERTURA DE DATOS EN COLOMBIA.” [Online]. Available: http://estrategia.gobiernoenlinea.gov.co/623/articles-9407_Guia_Apertura.pdf. [Accessed: 28-Jul-2015].
- [11] “LINEAMIENTOS PARA LA IMPLEMENTACIÓN DE DATOS ABIERTOS EN COLOMBIA.” [Online]. Available: https://www.fng.gov.co/ES/Documentos_DatosAbiertos/Modelo de Datos Abiertos.pdf. [Accessed: 05-May-2016].

Fecha de recepción: 3 de junio de 2016.

Fecha de aprobación: 9 de junio de 2016.

Walter Hugo Arboleda Mazo: warboleda@unac.edu.co

Jeovany Rojas Medina:
jerojas@unac.edu.co

Omer Urango:
ourango@unac.edu.co