

DOI: <https://doi.org/10.56712/latam.v5i2.1917>

Implementación de un portal de datos abiertos para la Universidad Ecuatoriana

Implementation of an open data portal for the Ecuadorian University

Gloria Arcos Medina

garcos@esepoch.edu.ec

<https://orcid.org/0000-0003-4601-0768>

Escuela Superior Politécnica de Chimborazo

Riobamba – Ecuador

Ivonne Rodríguez Flores

irodriguez@esepoch.edu.ec

<https://orcid.org/0000-0002-7788-2609>

Escuela Superior Politécnica de Chimborazo

Riobamba – Ecuador

Danilo Pástor Ramírez

dpastor@esepoch.edu.ec

<https://orcid.org/0000-0002-6144-3118>

Escuela Superior Politécnica de Chimborazo

Riobamba – Ecuador

Alejandra Oñate Andino

monate@esepoch.edu.ec

<https://orcid.org/0000-0001-7877-9198>

Escuela Superior Politécnica de Chimborazo

Riobamba – Ecuador

Marcelo Sevilla Vique

marcelosv86@gmail.com

Investigador Independiente

Riobamba – Ecuador

Artículo recibido: 19 de marzo de 2024. Aceptado para publicación: 02 de abril de 2024.

Conflictos de Interés: Ninguno que declarar.

Resumen

Actualmente, existen plataformas de código abierto que permiten la gestión y publicación de datos abiertos en un entorno web, siendo CKAN una de las más difundidas en organizaciones públicas y privadas. Sin embargo, existe limitada información para publicar y poner en marcha un portal de datos abiertos. El presente trabajo describe la implementación de un portal de datos abiertos para una universidad ecuatoriana utilizando la plataforma CKAN. Para apoyar este proceso se aplicó un procedimiento que consta de cinco fases, que se inicia con la instalación y personalización de CKAN, luego se establece las organizaciones y grupos necesarios, y culmina con la gestión de dataset y las configuraciones adicionales requeridas para garantizar la seguridad de los datos. Como resultado se obtuvo la plataforma CKAN instalada, configurada, personalizada y puesta en marcha para realizar la publicación del dataset dentro del portal de una universidad ecuatoriana. Producto del despliegue y personalización del portal de datos abiertos se crearon 41 organizaciones, 6 grupos temáticos y 32 dataset de los cuales 13 han sido publicados en esta primera fase del estudio. CKAN constituye una plataforma contemporánea y efectiva para publicar datos abiertos en instituciones gubernamentales educativas, pues cumple con los objetivos de publicar, compartir, editar, actualizar datos.

Palabras clave: CKAN, datos abiertos, dataset

Abstract

Currently, there are open-source platforms that allow the management and publication of open data in a web environment, with CKAN being one of the most widespread in public and private organizations. However, there is limited information to publish and launch an open data portal. This paper describes the implementation of an open data portal for an Ecuadorian university using the CKAN platform. To support this process, a procedure consisting of five phases was applied, which begins with the installation and customization of CKAN, then the necessary organizations and groups are established, and culminates with the management of dataset and the additional configurations required to guarantee security of the data. As a result, the CKAN platform installed, configured, customized and launched was obtained to publish the dataset within the portal of an Ecuadorian university. As a result of the deployment and customization of the open data portal, 41 organizations, 6 thematic groups and 32 datasets were created, of which 13 have been published in this first phase of the study. CKAN constitutes a contemporary and effective platform to publish open data in educational government institutions, as it meets the objectives of publishing, sharing, editing, updating data.

Keywords: CKAN, open data, dataset

Todo el contenido de LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades, publicados en este sitio está disponibles bajo Licencia Creative Commons . 

Cómo citar: Arcos Medina, G., Rodríguez Flores, I., Ramírez, D. P., Oñate Andino, A., & Sevilla Vique, M. (2024). Implementación de un portal de datos abiertos para la Universidad Ecuatoriana. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades* 5 (2), 807 – 823.
<https://doi.org/10.56712/latam.v5i2.1917>

INTRODUCCIÓN

Datos abiertos (Open Data, en inglés) es una práctica y a su vez una corriente ampliamente aceptada para publicar y compartir datos. Según la Carta Internacional de Datos Abiertos (Open Data Charter - ODC) los datos abiertos son “datos digitales que están disponibles con las características técnicas y legales necesarias para que cualquiera pueda utilizarlos, reutilizarlos y redistribuirlos libremente, en cualquier momento y en cualquier lugar” (Open Data Charter, 2015).

En América Latina mediante las tecnologías de la información se han desarrollado diversas iniciativas de datos abiertos, a través de portales web se ponen a disposición de la ciudadanía los datos para su reutilización y de esta manera agregar valor (Barenque, Belbis, & Serale, 2020; Naser & Rosales, 2016). En el contexto de Ecuador, en 2015 con tecnología CKAN se desarrolló el primer portal nacional de datos abiertos y en el año 2019 el país pasa a formar parte de la Alianza por el Gobierno Abierto (OGP por las siglas en inglés de Open Government Partnership). Mientras que, en 2021 el Gobierno como parte de la lucha contra la corrupción y la transparencia para el acceso a la información, a través del Ministerio de Telecomunicaciones y de la Sociedad de la Información (MINTEL) en cooperación con organizaciones internacionales como el Banco de Desarrollo de América Latina (CAF) y la Carta Internacional de Datos Abiertos (ODC) han trabajado para el lanzamiento de un nuevo portal nacional de datos abiertos (<https://datosabiertos.planificacion.gob.ec/>) para la publicación de conjunto de datos (dataset en inglés) con base en la Guía de Apertura de Datos Anti-corrupción de la ODC. A su vez, se realizó el relevamiento de datos publicados por los portales web oficiales del Ecuador de donde se evidencia que son pocas las entidades académicas que trabajan en datos abiertos (Open Data Charter, 2021).

De acuerdo con el trabajo de (Pástor, Rodríguez, Oñate-andino, & Arcos-medina, 2021), actualmente en Ecuador se ha identificado que únicamente cinco portales de datos abiertos corresponden a iniciativas que provienen de la Academia; de los cuales, un portal no está disponible y de los cuatro restantes, tres se relacionan con bibliotecas digitales y uno comparte conjuntos de datos geográficos. Por otro lado, si bien existen algunas investigaciones de datos abiertos en la universidad ecuatoriana, tales como (Labanda-Jaramillo et al., 2018; Vela Cadena & Enríquez Reyes, 2017); sin embargo, se ha evidenciado que algunos de los portales a los que hacen referencia en la actualidad no están vigentes.

CKAN (por las siglas en inglés de Comprehensive Knowledge Archive Network) es un sistema de gestión de datos abiertos (DMS) de código abierto que facilita la publicación, el intercambio y el uso de datos (CKAN, n.d.-a). El proyecto de código abierto CKAN es liderado por Open Knowledge Foundation (OKF) y en la actualidad utilizado para portales gubernamentales de datos abiertos (Máchová & Lněnička, 2017)(Oliveira, de Oliveira, Oliveira, & Lóscio, 2016). Esta plataforma permite la administración en dos entornos diferentes, el de publicación y el de desarrollo, los cuales tienen requerimientos tanto de hardware como software.

El objetivo del artículo es describir la implementación de un portal de datos abiertos para la universidad ecuatoriana utilizando CKAN, a través de un procedimiento de cinco pasos.

El resto del contenido del paper, está dividido en tres secciones: en la Sección II, se describe el marco teórico que fundamenta esta investigación, en la Sección III, se detalla la Metodología aplicada y cada una de sus fases, en la Sección IV se dan a conocer los resultados obtenidos del procedimiento aplicado y su discusión, finalmente en la sección V, se presentan las conclusiones del estudio.

DESARROLLO

En la actualidad, los datos abiertos (Open Data en inglés) es un término muy reconocido y está ampliamente difundido. De acuerdo con la Open Knowledge Foundation (Open Knowledge Foundation, n.d.) los datos abiertos son aquellos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, poseen derechos de atribución y se pueden compartir tal como se publican. Por otro lado, según el reporte titulado “Science as an open enterprise” (The Royal Society Science Policy Centre, 2012) para considerarse datos abiertos, se debe cumplir con las características que a continuación se indican:

Accesible: los datos deben ubicarse, de tal manera que se puedan encontrar fácilmente.

Evaluable: los destinatarios deben poder emitir algún juicio o evaluar lo que se comunica.

Inteligible: los datos proporcionan una descripción de los resultados del trabajo científico que sea inteligible para aquellos que deseen comprenderlos y analizarlos.

Usable: los datos deben poder reutilizarse, a menudo para diferentes propósitos.

Se pueden obtener muchos beneficios con la aplicación y uso de los datos abiertos. En el estudio de (Janssen, Charalabidis, & Zuiderwijk, 2012) se mencionan algunos de estos beneficios que pueden categorizarse como se indica en la Tabla 1.

Tabla 1

Beneficios de los datos abiertos por categoría

Categoría	Beneficio
Político y social	Más transparencia Responsabilidad democrática Más participación y auto-empoderamiento de los ciudadanos (usuarios) Creación de confianza en el gobierno Compromiso público Acceso igualitario a los datos
Económico	Crecimiento económico y estímulo de la competitividad Estimulación de la innovación Contribución a la mejora de procesos, productos y/o servicios Desarrollo de nuevos productos y servicios Uso de la sabiduría de las multitudes: aprovechar la inteligencia del colectivo
Operacional y técnico	La capacidad de reutilizar datos y contrarrestar la duplicación innecesaria Optimización de procesos administrativos Acceso a capacidad externa de resolución de problemas Toma de decisiones justa al permitir la comparación Acceso más fácil a los datos

Los datos abiertos cumplen un papel importante en la sociedad en general, es así como, las entidades del sector público se constituyen en las principales productoras de datos abiertos, a los cuales se los conoce como datos abiertos gubernamentales o por sus siglas en inglés OGD (Open Government Data). A los OGD se los consideran como una subcategoría de los datos abiertos que tienen como finalidad crear un gobierno más transparente, permitiendo que los ciudadanos estén más conscientes de los servicios prestados y obteniendo como resultado, una mejor relación entre el gobierno y los ciudadanos (Macêdo, 2018; Ubaldi, 2013).

Para que los datos estén disponibles para su uso es necesario publicarlos en portales web. El informe técnico de (Ham & Crockett, 2021) compara las principales plataformas comerciales y de código abierto de gestión y publicación de datos abiertos en un entorno web, tales como: CKAN

(Comprehensive Knowledge Archive Network), DKAN (Drupal-based open data portal based on CKAN), Socrata, EDX (Energy Data eXchange), Junar, entre otras. En el presente trabajo se utiliza CKAN por tener un soporte amplio de comunidades de apoyo, opciones de personalización y se adapta adecuadamente con los sistemas del gobierno.

CKAN es una plataforma de portal de datos basada en la web, desarrollada con el objetivo de proporcionar colecciones de datos para que otros publiquen, encuentren y usen datos abiertos en un entorno de software (Macêdo, 2018). Esta plataforma se caracteriza por manejar un backend construido en Python, y un frontend construido en JavaScript, también usa PostgreSQL como motor de base de datos, corre en el sistema operativo Linux y soporta el estándar DCAT (Data Catalog Vocabulary).

Una de las funcionalidades elementales de la plataforma CKAN es la catalogación de datos, la cual, se encarga de clasificar los conjuntos de datos por categorías y también ejecutar búsquedas específicas en el contenido del catálogo. A pesar de que los conjuntos de datos se referencian a través de enlaces, se puede configurar para almacenar los recursos directamente en CKAN (CKAN, n.d.-b). Para organizar los datos se utiliza un mecanismo de publicación llamado conjunto de datos (dataset), el cual, contiene información referente a un grupo de recursos. Por ejemplo, podría ser las estadísticas de repetición de una asignatura, la cantidad de carreras en un determinado período académico, cifras de estudiantes matriculados por carrera de una universidad, etc.

Cuando se realizan búsquedas en la plataforma CKAN se pueden desplegar un listado de conjunto de datos específicos, que pueden contener dos elementos (CKAN, n.d.-b; Herrera-Cubides, Gaona-García, & Gordillo Orjuela, 2017), que son:

Los metadatos. - proporcionan información como el título, descripción, etiquetas, licencia, propietario, fechas, entre otros.

Los recursos. - contiene los datos en sí, es prácticamente un archivo con tipo de formato específico como PDF, JSON, CSV, XLS, RDF, entre otros.

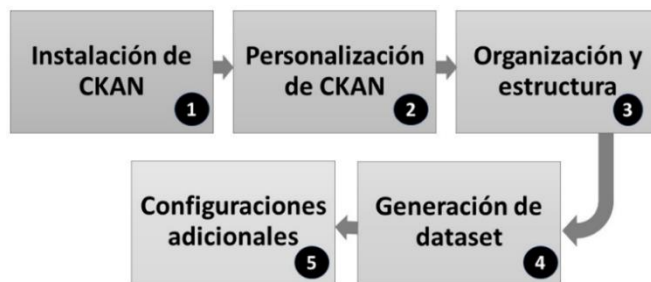
Para obtener los datos de sus fuentes originales y poder procesarlos adecuadamente como un conjunto de datos se debe seguir un proceso previo de extracción, transformación y carga denominado (ETL). Este proceso consiste en realizar determinadas actividades elementales (Vassiliadis, 2009), tales como: (i) extraer los datos de una o más fuentes; (ii) ejecutar la transformación y limpieza de los datos; y (iii) cargar los datos al destino apropiado. En la mayoría de los casos, se ejecutan las actividades descritas, añadiendo incluso tareas intermedias que se adaptan a la extracción de datos en distintos sistemas o bases de datos para poder cargarlos adecuadamente en una plataforma de datos abiertos. Se recomienda, crear scripts ETL para lograr obtener los datos abiertos refinados previo a la publicación en la plataforma CKAN.

METODOLOGÍA

Para el cumplimiento del objetivo de este paper, se aplicó un procedimiento que consta de cinco fases, que se inicia con la instalación de CKAN y culmina con la gestión de dataset y las configuraciones adicionales requeridas para garantizar la seguridad de los datos, que se presentan en la Figura 1.

Figura 1

Procedimiento para la implementación de CKAN



Fase 1: Instalación de CKAN

La instalación de CKAN se la puede realizar de tres maneras diferentes: 1) instalación en un sistema operativo desde un paquete, 2) instalación desde la fuente y 3) instalación empleando Docker Compose; cada una de las diferentes instalaciones permite cubrir un escenario y requerimientos diferentes. La instalación de CKAN desde el paquete se constituye en la manera más práctica de poner en marcha las diferentes funcionalidades de CKAN siempre y cuando se cuente con un equipo que tenga como sistema operativo Ubuntu en sus versiones 18.04 o 20.04 y cuando el objetivo no sea el de desarrollar o modificar la plataforma CKAN (CKAN, n.d.-b); por esta razón fue el método aplicado en esta investigación.

La instalación de CKAN demanda de varios paquetes que se constituyen en elementos básicos y que forman parte de la arquitectura de CKAN, en conjunto estos paquetes al ser descargados, instalados y configurados juntamente con los paquetes propios de CKAN dan origen a la plataforma de datos abiertos, los paquetes requeridos por CKAN se encuentran en la Tabla 2.

Tabla 2

Paquetes requeridos por CKAN

Paquete	Descripción
Python3	Entorno virtual de Python
PostgreSQL	DBMS para gestión de las bases de datos requeridas por CKAN
Apache Solr	Servidor empleado para búsquedas dentro de la plataforma
NGINX	Se lo emplea como proxy
uWSGI 8080/8800	Empleado como servidor web/empleado para el datapusher
Redis	Empleado para las búsquedas

Los paquetes que se encuentran en la tabla anterior deben estar instalados y configurados en el servidor en el que se desea instalar y desplegar la plataforma CKAN, cada uno de los paquetes posee una configuración específica. El proceso de instalación desde un paquete ha sido dividido en tres tareas: 1) instalación de herramientas software básicas, 2) instalación de PostgreSQL, 3) instalación de CKAN y 4) despliegue de CKAN.

La instalación de las herramientas software base para la instalación y despliegue de CKAN que se muestran en la tabla anterior requiere privilegios de administrador en el sistema operativo donde se hospedará el servidor de CKAN, para el presente caso Ubuntu 20.04, empleando los comandos apt install, wget y dpkg se realiza la instalación y configuración de las herramientas software. Es

indispensable tener en cuenta las versiones que se desean instalar, así como la compatibilidad del sistema operativo con la versión de las herramientas software.

Instalación y configuración de PostgreSQL, es necesario crear una base de datos, un usuario y la asignación de permisos requeridos al usuario que administra la base de datos creada que es utilizada CKAN.

La instalación de los paquetes CKAN requiere realizar previamente la descarga de los paquetes empleando el comando `wget` y direccionando directamente a la URL de descarga de los paquetes junto con la versión deseada para la descarga. Una vez finalizada la descarga se procede con la instalación de los paquetes de CKAN mismos que en la instalación desde el paquete requieren un mínimo de cambios a nivel de la configuración.

Para el despliegue de la plataforma CKAN, se deben reiniciar los servicios y servidores NGINX. Existen tres servicios, `ckan-datapusher`, `ckan-uwsgi`, `ckan-worker`, que se deben mostrar al momento de ejecutar el comando `supervisorctl status` que presenta la actividad, inactividad, éxito o fracaso de la puesta en marcha de la plataforma CKAN; si los tres servicios, se encuentran en ejecución la plataforma está instalada y funcional, bastará con un reinicio al servidor NGINX.

Fase 2: Personalización de CKAN

La fase de personalización de la plataforma depende del número de servicios, funcionalidades y cambios que se requiera a nivel de interfaz. En el archivo `ckan.ini`, constan todas las configuraciones que CKAN emplea al momento de desplegarse, entre las configuraciones que se encuentran en este archivo consta la URL, la configuración de notificaciones como las cuentas de correo desde las que se emiten las notificaciones, el correo en el que se recibe notificaciones en el caso de fallo o mal funcionamiento de la plataforma.

A nivel de la interfaz de usuario se pueden modificar ciertos archivos como páginas web y hojas de estilo en cascada para personalizar el diseño de la interfaz que posee la plataforma, por ejemplo: la cabecera, pie de página, colores, imágenes, etc.

Es importante habilitar complementos como: `datastore` y `datapusher`, ya que generan un cambio significativo a nivel de gestión y administración de la plataforma. La configuración del `datastore` demanda de la edición del archivo `ckan.ini`, mientras que la extensión `datapusher` requiere de la creación de un usuario y base de datos adicional en el que se almacenará la metadata de los archivos subidos a la plataforma. Además, el directorio en el que se guardan todos los archivos (`.jpg`, `.bmp`, `.xls`, `.csv`), debe ser configurado. La metadata correspondiente a estos archivos se almacena en la base de datos creada con este fin.

Una vez concluidos los procesos de instalación, configuración y personalización de la plataforma es el momento de desplegar la plataforma para verificar que las configuraciones surtan efecto, antes de ello se debe realizar un reinicio del servidor NGINX, `tomcat`. Luego, se debe comprobar el funcionamiento de los servicios configurados.

Fase 3: Organización y Estructura

Para la publicación de los dataset, previamente, es necesario establecer las organizaciones y grupos. Las organizaciones son elementos que permiten controlar la vista, creación y actualización de los dataset. Un dataset pertenece a una organización, la misma que controla el acceso a sus dataset ya sea de forma pública o privada.

En CKAN se publican dataset (conjuntos de datos), por ejemplo, los graduados por carreras por período académico. Los resultados de una búsqueda serán dataset individuales. Una instancia de CKAN puede incluir diferentes organizaciones. Por ejemplo, las unidades académicas de una Institución de educación superior, y cada de ellas publican los dataset (CKAN, n.d.-a).

CKAN provee una gestión de tres niveles de usuarios: Administrador, Editor, Miembro, a los que se incrementa el usuario Administrador de la Plataforma. Estos niveles de usuarios cuentan con permisos inferiores al del usuario Administrador de la Plataforma, pues el usuario Administrador de esta categoría únicamente tiene privilegios de administrador a nivel de una organización o de un grupo, de igual manera el usuario Editor tiene permisos para editar los dataset únicamente a nivel de una Organización o de un grupo, finalmente el usuario Miembro es un usuario que únicamente tiene permisos para visualizar las publicaciones.

Las organizaciones cuentan con los mismos niveles de usuarios (Administrador, Editor, Miembro), cada uno de ellos con funciones y permisos que van de acuerdo a su perfil. Los cambios, configuraciones, subida de archivos realizada por un usuario generan logs que permiten determinar quién realizó una determinada acción.

Los grupos dentro de la plataforma CKAN se los utiliza para crear y administrar un conjunto de dataset, que surgen en base a las necesidades de un determinado proceso, plan o proyecto; de esta manera, se puede contar con un conjunto de usuarios de diferentes organizaciones categorizados en un mismo grupo.

Fase 4: Generación de dataset

La generación de dataset que se publican en el CKAN se apoya de procesos de extracción, transformación y carga (ETL), para lo cual se realizaron los pasos que a continuación se listan:

Construir procesos ETLs que permitan: i) Acceder a las bases de datos académicas, como origen de datos, y extraer la información de interés que está relacionada con los indicadores académicos; ii) Realizar la limpieza de datos; y, iii) Cargar los datos extraídos y limpios a una base de datos intermedia relacional, la cual tiene por objetivo almacenar resúmenes de datos. Esta base de datos intermedia está esquematizada como estrella con un enfoque OLAP (por las siglas en inglés de On-Line Analytical).

Mediante scripts SQL, consultar la base de datos intermedia para extraer datos que se exportan a archivos con formatos abiertos, tales como .csv, .ods, entre otros. Y, estos archivos pasan a ser los recursos del dataset que se publica.

En la plataforma CKAN, subir el dataset y su documentación mediante la metadata correspondiente.

Fase 5: Configuraciones adicionales

Un elemento indispensable a nivel de todo servidor web es el contar con elementos de seguridad, estos elementos pueden ser: de seguridad activa (credenciales, permisos, software antivirus, actualizaciones, backups) y de seguridad pasiva (escaneo, limpieza de amenazas, análisis forense). Otro elemento a considerar son los elementos de seguridad a nivel hardware (cortafuegos, firewall, servidor proxy) y software (software antivirus, filtros anti spam, cortafuegos UFW) (Sepúlveda, Salcedo, & Gómez, 2012). Adicionalmente, es necesario instalar los certificados SSL, para la implementación del protocolo https a nivel de la plataforma CKAN.

RESULTADOS Y DISCUSIÓN

En este artículo se presentan los resultados de la aplicación de un procedimiento para la implementación de datos abiertos en la universidad ecuatoriana, se han generado dataset relacionados con la actividad académica a través de procesos de extracción, transformación y carga (ETL), garantizando que los datos hayan sido adecuadamente filtrados, depurados y que sean publicables. A continuación, se presentan los principales resultados:

Instalación de CKAN

Como se puede evidenciar en la Figura 2 se realiza el proceso de instalación del CKAN, para ello se instaló las herramientas software básicas, por ejemplo, Ubuntu 20.4; además PostgreSQL, y CKAN, logrando implementar y desplegar la plataforma como se puede evidenciar en la Figura 3.

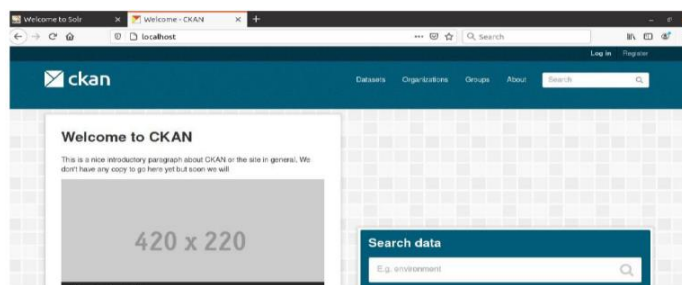
Figura 2

Interfaz de verificación de los servicios requeridos para el servidor CKAN

```
ckanuser@ckanepoch:~$ sudo supervisorctl reload
Restarted supervisord
ckanuser@ckanepoch:~$ sudo supervisorctl status
ckan-datapusher:ckan-datapusher-00    RUNNING   pid 13111, uptime 0:00:17
ckan-uwsgi:ckan-uwsgi-00              RUNNING   pid 13112, uptime 0:00:17
ckan-worker:ckan-worker-00           RUNNING   pid 13113, uptime 0:00:17
ckanuser@ckanepoch:~$ sudo service nginx restart
[ OK ] Stopping A high performance web server and a reverse proxy server...
[ OK ] Starting A high performance web server and a reverse proxy server...
```

Figura 3

Interfaz de inicio del servidor CKAN en el browser

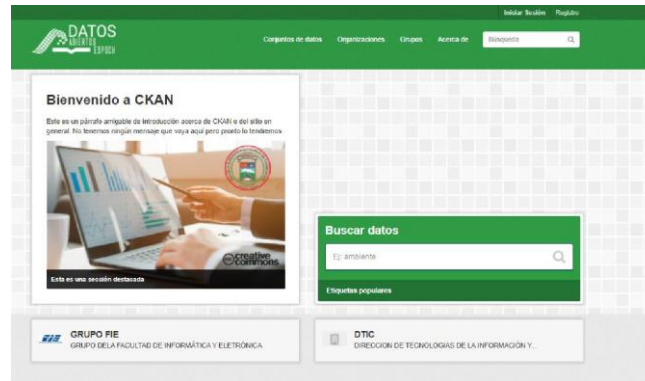


Personalización de CKAN

Una vez instalada la plataforma CKAN, se procede a personalizar el diseño de la interfaz, y se habilitan los complementos requeridos, teniendo como resultado la plataforma CKAN implementada y funcional. En la Figura 4 se muestra la interfaz principal de acceso al portal, desde la que es posible realizar cambios en el idioma, si es requerido, además, consta de todos los idiomas disponibles en CKAN activados y configurados para de esta manera ampliar el acceso a los datos abiertos.

Figura 4

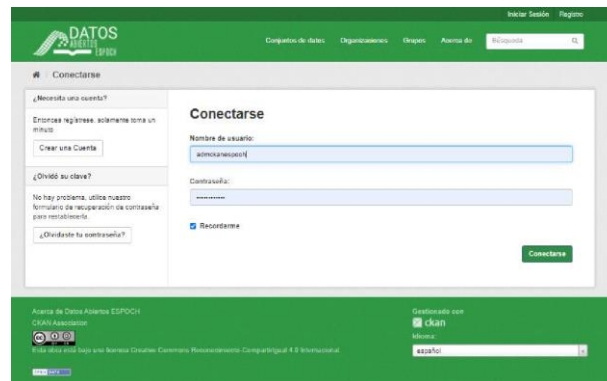
Interfaz principal CKAN



La siguiente interfaz corresponde al login requerido por la plataforma para de esta manera tener acceso a los recursos con base en los privilegios asignados para cada rol "Administrador, Editor, Miembro".

Figura 5

Interfaz de logueo CKAN



Estructura del Portal de Datos Abiertos Propuesta

Para la definición de la estructura del portal de datos abiertos propuesta, se consideró la estructura orgánica funcional de la universidad, la misma considera facultades, carreras, institutos, departamentos, por lo que en este estudio se han creado 41 organizaciones (ver figura 6).

Figura 6

Organizaciones creadas



Además, se han definido 6 grupos temáticos, establecidos en función de la información que se va a publicar, y que es requerida para dar cumplimiento a los diferentes indicadores de calidad, estos son:

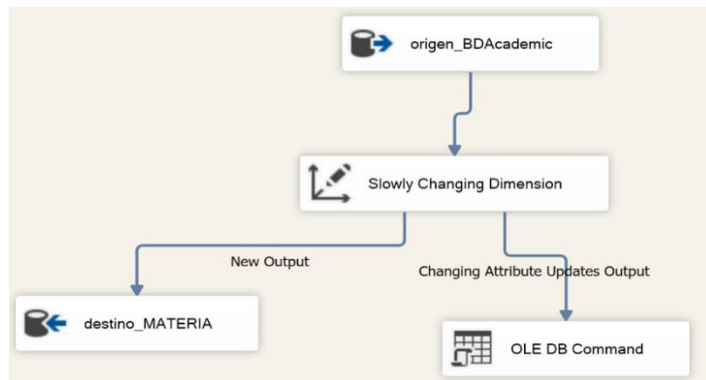
- Estudiantes.
- Información curricular.
- Profesores.
- Investigación y Vinculación.
- Administración.
- Infraestructura y servicios.

Generación de dataset

El movimiento de los datos desde un origen que son las bases de datos operacionales de la Universidad hacia un destino que, para este caso, es la base de datos intermedia se realiza con procesos ETLs (Ver Figura 7). Y, a su vez, esta base de datos intermedia es el origen para generar dataset que se publican en la plataforma CKAN. Dado el esquema de esta base de datos, también puede ser utilizada con herramientas de Inteligencia de Negocio.

Figura 7

Implementación ETL



Con el mecanismo descrito, se han generado 32 dataset de los cuales 13 han sido publicados en esta primera fase del estudio. En la Tabla 3 se detalla un ejemplo de la información que estos dataset contienen.

Tabla 3

Ejemplo de dataset generados en el grupo Estudiantes

ORGANIZACIÓN: Dirección de Tecnologías de la Información y Comunicación	
Grupo temático	DATASET
Estudiantes	Tasa de Titulación por carrera por período académico Tasa de Retención por carrera por período académico Rendimiento Académico por materia por carrera por período académico No. de graduados por carrera por carrera y por año académico (hombres, mujeres) No. Estudiantes por carrera por período académico (hombres, mujeres) No. de estudiantes por semestre por carrera por período académico académico (hombres, mujeres). No. de Estudiantes admitidos por carrera por período académico académico (hombres, mujeres). No. de becarios por carrera por período académico académico (hombres, mujeres). No. de discapacitados por carrera por período académico académico (hombres, mujeres). Porcentaje de alumnos aprobados por asignatura, por carrera o por período académico (hombres, mujeres).

Cuando se realizan búsquedas dentro del portal de datos abiertos propuesto, se muestra un listado de los dataset específicos, los mismos que contienen los recursos (Figura 8) y los metadatos correspondientes (Figura 9).

Figura 8

Dataset Portal Datos Abiertos ESPOCH

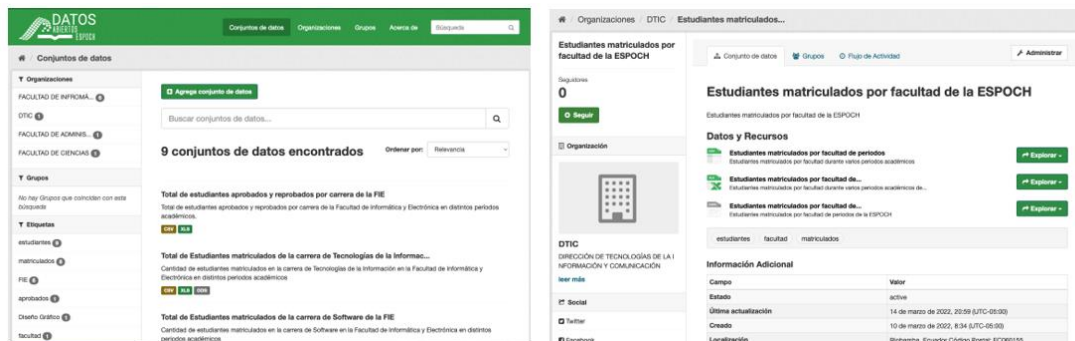


Figura 9

Metadata de estudiantes matriculados por facultad de la ESPOCH

Campos	Valor
Última actualización de los datos	10 de marzo de 2022
Última actualización de los metadatos	10 de marzo de 2022
Creado	10 de marzo de 2022
Formato	CSV
Licencia	Creative Commons Attribution Share-Alike
Has views	True
Id	dbd79e65-1be2-497c-b506-2a4013730014
Mimetype	text/csv
On same domain	True
Package id	e9592d38-86f9-4d7c-ac57-61ae1e3d190f
Size	2 KIB
State	active
Url type	upload

Configuraciones adicionales

Dentro de este proceso se realizó configuraciones que permitieron contar con elementos de seguridad para el acceso al portal de datos abiertos propuesto, se realizó la instalación y configuración de seguridades a nivel del servidor como por ejemplo el firewall (Figura 10). Además, se realizó la activación de certificados SSL, para la implementación del protocolo https a nivel de la plataforma CKAN (Ver Figura 11).

Figura 10

Interfaz de administrador del firewall

```
ckanuser@ckanespoch:~$ sudo ufw status
Status: active

To Action From
--
20/tcp ALLOW Anywhere
21/tcp ALLOW Anywhere
990/tcp ALLOW Anywhere
40000:50000/tcp ALLOW Anywhere
20/tcp (v6) ALLOW Anywhere (v6)
21/tcp (v6) ALLOW Anywhere (v6)
990/tcp (v6) ALLOW Anywhere (v6)
40000:50000/tcp (v6) ALLOW Anywhere (v6)
```

Figura 11

Pantalla de generación del certificado y la clave ssl

```
ckanespoch@ubuntu:~$ openssl genrsa 2048 > host.key
Generating RSA private key, 2048 bit long modulus (2 primes)
.....+++++
.....+++++
e 15 65537 (0x010001)
ckanespoch@ubuntu:~$ █

ckanespoch@ubuntu:~$ chmod 400 host.key
ckanespoch@ubuntu:~$ █
```

Una vez instalado el portal de datos abiertos, es importante resaltar algunos aspectos importantes que tuvieron que tomarse en cuenta para contar con una plataforma que publique dataset con un adecuado nivel de reutilización y apertura, entre estos aspectos podemos mencionar los siguientes:

Es necesario establecer los derechos de propiedad intelectual para los dataset publicados, de acuerdo con (Vicente-Saez, Gustafsson, & Van den Brande, 2020), así como la protección de información sensible de los datos de los participantes en los procesos universitarios y la creación de canales de comunicación suficientes y adecuados entre organizaciones para facilitar la publicación de la información (Borgerud & Borglund, 2020); estas dificultades encontradas en la ejecución de este estudio, coinciden con lo afirmado por Vicente-Saez y Borgerud, y han sido resueltas para viabilizar la implementación de datos abiertos.

Debe existir la voluntad de compartir datos por parte de los responsables de las universidades (Perkmann & Schildt, 2015), con el fin de que estos datos sean reutilizados en varios contextos que la academia lo requiera.

Acorde con (Vicente-Saez et al., 2020) es necesario definir una estandarización sobre el gobierno de datos, infraestructura, prácticas, protocolos de publicación, soporte técnico, etc.

Ha sido necesario estandarizar la metadata (Borgerud & Borglund, 2020), de acuerdo con lo exigido por CKAN y cumplir con los parámetros mínimos de apertura de datos.

El origen de los dataset considerados cuentan con diferentes principios de modelado, diseños y vocabularios lo que se constituyó en un problema al generar los dataset, tal como lo afirma (d'Aquin, 2016), que es una actividad que aún debe fortalecerse.

Existe dificultad en actualizar la información en tiempo real, como lo indica (Piros, Fleiner, & Kovács, 2017) ya que gran cantidad de información académica se genera semestralmente; ha sido necesario un proceso de depuración y sistematización de los datos antes de su publicación.

CONCLUSIONES

Las plataformas de datos abiertos permiten generar una interacción directa e inclusiva entre el gobierno y sus mandantes, entidades públicas y privadas con la sociedad en general. La universidad ecuatoriana cumple un papel preponderante en lo referente a gobierno abierto pues los temas académicos, de investigación, biblioteca, archivo no pueden estar excluidos en un gobierno abierto.

La elección de herramientas para la publicación de datos en portales de datos abiertos está en directa dependencia de los objetivos que se persigue con la publicación ya sean estos de desarrollo de aplicaciones, extensiones, o simplemente sean objetivos de publicación, distribución, edición y actualización de datos. CKAN se constituye en una eficaz y eficiente herramienta para publicación de

datos abiertos, pues cumple con los objetivos de publicar, compartir, editar, actualizar datos, además, provee herramientas de administración y gestión de usuarios, documentos, permisos, elementos que brindan eficiencia en las tareas de administración y gestión del portal.

Existen varios inconvenientes que limitan la implementación de datos abiertos en la universidad ecuatoriana, entre los más importantes se pueden mencionar los siguientes: la falta de estándares, soporte técnico, existencia de soluciones particulares que no se adaptan a situaciones generales, etc, otras dificultades encontradas están relacionados con la propiedad intelectual, aspectos legales y percepción de datos abiertos por parte de los involucrados.

El trabajo futuro está dirigido a mejorar la utilidad y reutilización de los datos abiertos, para lo cual en primera instancia será necesario llevar a cabo una evaluación del portal de datos abiertos propuesto.

REFERENCIAS

- Barenque, A., Belbis, J. I., & Serale, F. (2020). Implementación de estrategias de datos abiertos en América Latina. Retrieved from <https://idatosabiertos.org/publicaciones/estrategias-de-datos-abiertos-en-america-latina/>
- Borgerud, C., & Borglund, E. (2020). Open research data, an archival challenge? *Archival Science*, 1–24. <https://doi.org/https://doi.org/10.1007/s10502-020-09330-3>
- CKAN. (n.d.-a). CKAN - The open source data management system. Retrieved January 20, 2022, from <https://ckan.org/>
- CKAN. (n.d.-b). Installing CKAN from package – CKAN 2.9.3 documentation. Retrieved October 30, 2021, from <https://docs.ckan.org/en/2.9/maintaining/installing/install-from-package.html>
- d'Aquin, M. (2016). On the use of linked open data in education: Current and future practices. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9500, 3–15. https://doi.org/10.1007/978-3-319-30493-9_1
- Ham, K. D., & Crockett, D. L. (2021). Evaluation of Data Catalog Software for Hanford Site Environmental Datasets. <https://doi.org/10.2172/1832173>
- Herrera-Cubides, J. F., Gaona-García, P. A., & Gordillo Orjuela, K. (2017). Una Mirada a la Web de los Datos. Caso de Estudio: Consumo de Servicios CKAN. *Ingeniería*, 22(1), 46–64. <https://doi.org/10.14483/UDISTRITAL.JOUR.REVING.2017.1.A07>
- Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, 29(4), 258–268. <https://doi.org/10.1080/10580530.2012.716740>
- Labanda-Jaramillo, M., Chamba-Eras, L., Coronel-Romero, E., Granda, J.-L., Riofrío, A. M. G., & Zhingre, L. E. C. (2018). Open Data Portal Prototype for the Conservation of Biodiversity of the 'Universidad Nacional de Loja.' *KnE Engineering*, 3(9), 133. <https://doi.org/10.18502/KEG.V3I9.3651>
- Macêdo, J. J. de. (2018). OpenEasier: a CKAN extension to enhance open data publication and management (Universidade Federal do Rio Grande do Norte). Retrieved from <https://repositorio.ufrn.br/jspui/handle/123456789/26025>
- Máchová, R., & Lněnička, M. (2017). Evaluating the Quality of Open Data Portals on the National Level. *Journal of Theoretical and Applied Electronic Commerce Research* 2017, Vol. 12, Pages 21-41, 12(1), 21–41. <https://doi.org/10.4067/S0718-18762017000100003>
- Naser, A., & Rosales, D. (2016). Panorama regional de los datos abiertos: avances y desafíos en América Latina y el Caribe. *Serie Gestión Pública*, 86. Retrieved from <https://repositorio.cepal.org/handle/11362/40768>
- Oliveira, M. I. S., de Oliveira, H. R., Oliveira, L. A., & Lóscio, B. F. (2016). Open Government Data Portals Analysis: The Brazilian Case. *Proceedings of the 17th International Digital Government Research Conference on Digital Government Research*, 415–424. <https://doi.org/10.1145/2912160.2912163>
- Open Data Charter. (2015). International Open Data Charter. Principles. Retrieved April 20, 2021, from <https://opendatacharter.net/principles/>

Open Data Charter. (2021). Abriendo datos anticorrupción en Ecuador. Retrieved January 19, 2022, from <https://medium.com/opendatacharter/abriendo-datos-anticorrupción-en-ecuador-8d690d21906>

Open Knowledge Foundation. (n.d.). The Open Data Handbook. Retrieved February 14, 2022, from <http://opendatahandbook.org/>

Pástor, D., Rodríguez, I. E., Oñate-andino, A., & Arcos-medina, G. (2021). REUSING OPEN DATA: AN EXPLORATORY STUDY IN ECUADOR. *Journal of Theoretical and Applied Information Technology*, 15(13). Retrieved from www.jatit.org

Perkmann, M., & Schildt, H. (2015). Open data partnerships between firms and universities: The role of boundary organizations. *Research Policy*, 44(5), 1133–1143. <https://doi.org/10.1016/j.respol.2014.12.006>

Piros, P., Fleiner, R., & Kovács, L. (2017). Linked data generation for courses and events at Óbuda University. 2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI), 253–258. <https://doi.org/10.1109/SAMI.2017.7880313>

Sepúlveda, E., Salcedo, O., & Gómez, E. (2012). MANEJO DEL RIESGO Y SEGURIDAD EN EL CONSUMO DE SERVICIOS DE TI EN CLOUD COMPUTING. *Redes de Ingeniería*; Vol. 1 Núm. 2 (2010); 10-21, 1(2), 10. <https://doi.org/10.14483/2248762X.7156>

The Royal Society Science Policy Centre. (2012). Science as an open enterprise. Retrieved from https://royalsociety.org/~media/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf

Ubaldi, B. (2013). Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives. In *OECD Working Papers on Public Governance*. <https://doi.org/https://doi.org/10.1787/5k46bj4f03s7-en>

Vassiliadis, P. (2009). A Survey of Extract-Transform-Load Technology. *International Journal of Data Warehousing and Mining (IJDWM)*, 5(3), 1–27. <https://doi.org/10.4018/jdwm.2009070101>

Vela Cadena, S. G., & Enríquez Reyes, R. A. (2017). Portal de datos abiertos de la Universidad Central del Ecuador. *FIGEMPA: Investigación y Desarrollo*, 1(2), 85–90. <https://doi.org/10.29166/REVFIG.V1I2.73>

Vicente-Saez, R., Gustafsson, R., & Van den Brande, L. (2020). The dawn of an open exploration era: Emergent principles and practices of open science and innovation of university research teams in a digital world. *Technological Forecasting and Social Change*, 156, 120037. <https://doi.org/10.1016/j.techfore.2020.120037>